

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 3, 2003, 19:47:58 ; Search time 15.5 Seconds
(without alignments)
94.913 Million cell updates/sec

Title: US-09-725-010-2

Perfect score: 279

Sequence: 1 CTNCGTKTTPLRNPPGQP.....LHGVRPLSLKTDVKKRQR 50

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- Issued_Patents_AA.*
- 1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep.*
- 2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep.*
- 3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep.*
- 4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep.*
- 5: /cgn2_6/ptodata/1/iaa/PTUS_COMB.pep.*
- 6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	258	92.5	853	3	US-09-254-325-2
2	258	92.5	866	3	US-09-079-415-6
3	258	92.5	866	3	US-08-750-458A-2
4	188	67.4	449	2	US-08-927-394-2
5	54	19.4	108	1	US-08-204-740-11
6	54	19.4	108	3	US-09-081-167A-11
7	54	19.4	108	3	US-09-081-393-11
8	54	19.4	108	4	US-09-416-833-11
9	54	19.4	108	5	PCR-US95-02521-11
10	51.5	18.5	341	4	US-09-008-465-1
11	51.5	18.5	341	4	US-09-528-959-1
12	51	18.3	416	1	US-08-117-083-61
13	51	18.3	543	4	US-09-362-123A-4
14	50.5	18.1	314	1	US-08-269-441A-2
15	50.5	18.1	314	4	US-09-008-465-3
16	50.5	18.1	314	4	US-09-276-851-2
17	50.5	18.1	314	4	US-09-528-959-3
18	50.5	18.1	381	3	US-08-911-853-27
19	50.5	18.1	381	4	US-09-479-409-27
20	50.5	18.1	381	4	US-09-479-453-27
21	50	17.9	664	2	US-08-852-153-8
22	50	17.9	1068	3	US-08-390-874C-11
23	50	17.9	1068	2	US-09-265-772-11
24	50	17.9	1069	2	US-08-162-081B-37
25	50	17.9	1069	2	US-08-780-872-37
26	50	17.9	1069	2	US-09-085-957-37
27	50	17.9	1080	2	US-08-162-081B-36

28	50	17.9	1080	2	US-08-780-872-36	Sequence 36, Appl
29	50	17.9	1080	4	US-09-085-957-36	Sequence 36, Appl
30	50	17.9	2372	4	US-09-579-181-2	Sequence 2, Appl
31	50	17.9	3118	4	US-09-579-181-1	Sequence 1, Appl
32	49.5	17.7	249	4	US-09-134-001C-4442	Sequence 4442, Ap
33	49.5	17.7	657	4	US-09-355-166-1	Sequence 1, Appl
34	49	17.6	128	5	PCT-US95-06266-141	Sequence 141, App
35	49	17.6	191	3	US-08-974-022-52	Sequence 52, Appl
36	49	17.6	191	4	US-08-795-445A-52	Sequence 52, Appl
37	49	17.6	191	4	US-08-795-447A-52	Sequence 52, Appl
38	49	17.6	191	4	US-08-974-186-52	Sequence 52, Appl
39	49	17.6	191	4	US-08-795-446B-52	Sequence 52, Appl
40	49	17.6	191	4	US-08-706-945D-139	Sequence 139, App
41	49	17.6	256	1	US-08-236-918A-6	Sequence 6, Appl
42	49	17.6	256	4	US-09-150-864A-6	Sequence 6, Appl
43	49	17.6	256	4	US-08-012-269A-2	Sequence 2, Appl
44	49	17.6	256	5	PCT-US96-03965-2	Sequence 2, Appl
45	49	17.6	2910	1	US-08-466-033-183	Sequence 183, App

ALIGNMENTS

RESULT 1
US-09-254-325-2
; Sequence 2, Application US/09254325
; Patent No. 6090607
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: ENHANCED EXPRESSION OF
; TITLE OF INVENTION: PROTEOLYTIC ENZYMES IN KOJI MOLDS
; NUMBER OF SEQUENCES: 17
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/254,325
; FILING DATE:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 853 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FEATURE:
; NAME/KEY: Binding-site
; LOCATION: 652-676
; OTHER INFORMATION: /note= "DNA BINDING SITE"
; FEATURE:
; NAME/KEY: Region
; LOCATION: 1..731
; OTHER INFORMATION: /note= "TRUNCATED AREA WHICH IS
; OTHER INFORMATION: STILL ACTIVE BUT NOT RERESSED BY
; OTHER INFORMATION: L-GLUTAM..."
US-09-254-325-2

Query Match 92.5%; Score 258; DB 3; Length 853;
Best Local Similarity 92.0%; Pred. No. 8.6e-29;
Matches 46; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 CTNCGTKTTPLRNPPGQPLCNACGLFLKLGHVVRPLSLKTDVKKRQR 50
||||| :||||| :||||| :||||| :||||| :||||| :||||| :
Db 652 CTNCFQTTPLRNPPGQPLCNACGLFLKLGHVVRPLSLKTDVKKRNR 701

RESULT 2
US-09-079-415-6
; Sequence 6, Application US/09079415
; Patent No. 6013452
; GENERAL INFORMATION:

```
; REFERENCE/DOCKET NUMBER: 4129,204-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 867 0123
; TELEFAX: 212 867 0298
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 866 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-750-458A-2
Query Match          92.5%; Score 258; DB 3; Length 866;
Best Local Similarity 92.0%; Pred. No. 8.7e-29;
Matches 46; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY      1 CTNCGTGTTPLWRRNPQGQPLCNACGLFLKHLGVVRLSLKTDVIKKRQR 50
        ||||| :||||| :||||| :||||| :||||| :||||| :||||| :
Db       664 CTCNCFQTPTLWRRNPEGOPLCNACGLFLKHLGVVRLSLKTDVIKKRN 713

RESULT 4
US-08-927-394-2
; Sequence 2, Application US/08927394
; Patent No. 5990092
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: GATA-6 TRANSCRIPTION FACTOR: COMPOSITIONS AND METHODS
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wolf, Greenfield, & Sacks, P.C.
; STREET: 600 Atlantic Avenue
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/927,394
; FILING DATE:
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/025,574
; FILING DATE: 06-SEP-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Plumer, Elizabeth R.
; REGISTRATION NUMBER: 36,637
; REFERENCE/DOCKET NUMBER: S1237/7005
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-720-3500
; TELEFAX: 617-720-2441
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 449 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-927-394-2
Query Match          67.4%; Score 188; DB 2; Length 449;
Best Local Similarity 62.0%; Pred. No. 5e-19;
Matches 31; Conservative 10; Mismatches 9; Indels 0; Gaps 0;

QY      1 CTNCGTGTTPLWRRNPQGQPLCNACGLFLKHLGVVRLSLKTDVIKKRQR 50
        ||||| :||||| :||||| :||||| :||||| :||||| :||||| :
Db       298 CANGCHTTTTLWRRNAEGEPCVNACGLYMKLHGVRPLAMKKEGIQTRKR 347

RESULT 5
```


RESULT 2

Q9Y7E8 PRELIMINARY; PRT; 866 AA.
AC Q9Y7E8;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE Major nitrogen regulatory protein.
GN AREA.
OS Aspergillus parasiticus.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
OX NCBI_TaxID=5067;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 56775;
RX MEDLINE=20225467; PubMed=10760588;
RA Chang P.-K., Yu J., Bhatnagar D., Cleveland T.E.;
RT "Characterization of Aspergillus parasiticus major nitrogen regulatory
gene, area."
RL Biochim. Biophys. Acta 1491:263-266(2000).
DR EMBL; AF148539; AAD37409.1; -.
DR HSSP; P17429; 4GAT.
DR InterPro; IPR000679; Znf_GATA.
DR PRINTS; PR00619; GATAZNFINGER.
DR SMART; SM00401; Znf_GATA; 1.
DR PROSITE; PS00344; GATA_ZN_FINGER_1; 1.
DR PROSITE; PS01114; GATA_ZN_FINGER_2; 1.
SQ SEQUENCE 866 AA; 92877 MW; 16FF4D56E823988 CRC64;

Query Match 92.5%; Score 258; DB 3; Length 866;
Best Local Similarity 92.0%; Pred. No. 2.1e-26;
Matches 46; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
QY 1 CTNCGTKTTPLRNPOGQPLCNACGLFLKLGWVRPLSLKTDVKKRQ 50
Db 664 CTNCFQTTPLRNPEGQPLCNACGLFLKLGWVRPLSLKTDVKKRNR 713

RESULT 3

Q9HEW7 PRELIMINARY; PRT; 918 AA.
AC Q9HEW7;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Nitrogen response factor NRFL.
OS Cladosporium fulvum (Fulvia fulva).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina;
OC Dothideomycetes et Chaetothyriomycetes incertae sedis;
OC Mycosphaerellaceae; mitosporic Mycosphaerellaceae; Cladosporium.
OX NCBI_TaxID=5499;
RN [1]
RP SEQUENCE FROM N.A.
RA Perez-Garcia A., Snoeijers S.S., Joosten M.H.A.J., Goosen T.,
De Wit P.J.G.M.;
RT "Expression of the avirulence gene Avr9 of the fungal tomato pathogen
Cladosporium fulvum is regulated by the global nitrogen response
factor NRFL."
RL Mol. Plant Microbe Interact. 0:0-0(2001).
DR EMBL; AF312694; AAG48616.1; -.
DR HSSP; P17429; 4GAT.
DR InterPro; IPR002965; P_rich_extensn.
DR InterPro; IPR000679; Znf_GATA.
DR Pfam; PF00320; GATA; 1.
DR PRINTS; PR00619; GATAZNFINGER.
DR PROSITE; PS00344; GATA_ZN_FINGER_1; 1.
DR PROSITE; PS01114; GATA_ZN_FINGER_2; 1.
SQ SEQUENCE 918 AA; 99077 MW; E98198D999BCA899 CRC64;

Query Match 92.5%; Score 258; DB 3; Length 918;
Best Local Similarity 92.0%; Pred. No. 2.2e-26;
Matches 46; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 CTNCGTKTTPLRNPOGQPLCNACGLFLKLGWVRPLSLKTDVKKRQ 50
Db 673 CTNCFQTTPLRNPEGQPLCNACGLFLKLGWVRPLSLKTDVKKRNR 722

RESULT 4

O60043 PRELIMINARY; PRT; 944 AA.
AC O60043;
DT 01-AUG-1998 (TREMBLrel. 07, Created)
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE Nitrogen response regulator.
GN NRRL.
OS Metarhizium anisopliae.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Hypocreales; Clavicipitaceae; mitosporic Clavicipitaceae; Metarhizium.
OX NCBI_TaxID=5530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MEI;
RX MEDLINE=99070047; PubMed=9852945;
RA Screen S.E., Bailey A.M., Charnley K., Cooper R., Clarkson J.M.;
RT "Isolation of a nitrogen response regulator gene (nr1) from
Metarhizium anisopliae."
RL Gene 221:17-24(1998).
DR EMBL; AJ006468; CAA07052.1; -.
DR HSSP; P17429; 4GAT.
DR InterPro; IPR000679; Znf_GATA.
DR Pfam; PF00320; GATA; 1.
DR PRINTS; PR00619; GATAZNFINGER.
DR SMART; SM00401; Znf_GATA; 1.
DR PROSITE; PS00344; GATA_ZN_FINGER_1; 1.
DR PROSITE; PS01114; GATA_ZN_FINGER_2; 1.
SQ SEQUENCE 944 AA; 99687 MW; A1A723B5658C23EAO CRC64;

Query Match 92.5%; Score 258; DB 3; Length 944;
Best Local Similarity 92.0%; Pred. No. 2.3e-26;
Matches 46; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 CTNCGTKTTPLRNPOGQPLCNACGLFLKLGWVRPLSLKTDVKKRQ 50
Db 663 CTNCFQTTPLRNPEGQPLCNACGLFLKLGWVRPLSLKTDVKKRNR 712

RESULT 5

P87016 PRELIMINARY; PRT; 327 AA.
AC P87016;
DT 01-JUL-1997 (TREMBLrel. 04, Created)
DT 01-JUL-1997 (TREMBLrel. 04, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE DNA-binding protein.
GN SAT1.
OS Zygosaccharomyces rouxii (Candida mogii).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Zygosaccharomycetes.
OX NCBI_TaxID=4956;
RN [1]
RP SEQUENCE FROM N.A.
RA Ushio K., Otsuka H., Yoshikawa S., Taguchi G., Shimosaka M.,
Mitsui N., Okazaki M.;
RT "Cloning of the SAR1 gene concerned with salt tolerance of the yeast
Zygosaccharomyces rouxii."
RL J. Ferment. Bioeng. 82:16-21(1996).
DR EMBL; D83211; BAA18948.1; -.
DR HSSP; P17429; 4GAT.
DR InterPro; IPR000679; Znf_GATA.
DR Pfam; PF00320; GATA; 1.


```

OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA Champe R., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,
RA George M., Gonzalez M., Guarin H., Kronmiller B., Li P., Liao G.,
RA Miranda A., Mungall C.J., Nunoo J., Pacleb J., Paragas Y., Park S.,
RA Patel S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
RA Ceiniker S.;
RL Submitted (APR-2002) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AY095095; AACM11423.1; -.
SQ SEQUENCE 486 AA; 51606 MW; BC0E10E9B8D40A0B CRC64;

Query Match 71.0%; Score 198; DB 5; Length 486;
Best Local Similarity 66.0%; Pred. No. 1.6e-18;
Matches 33; Conservative 9; Mismatches 8; Indels 0; Gaps

Qy 1 CTNCGKTTPLWRNPQGOPLCNACGLFLKLHGVRPLSLKTDVKKQR 50
|||||:||||| |:|:|||||: ||||| |:|:| | :| :| :|
Db 174 CTNCGTTRTTLWRRNDGEPVCNAGLYXKYLHGVNRPLAMRKDGIOTKR 223
|||||:||||| |:|:|||||: ||||| |:|:| | :| :| :|

RESULT 15
Q9W6UO PRELIMINARY; PRT; 383 AA.
ID AC Q9W6UO
IC Q9W600;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DE 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Gata5.
GN GATA5 OR GTA5.
OS Brachydanio rerio (Zebrafish) (zebra danio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=EMBRYO;
RX MEDLINE=99307072; PubMed=10375499;
RA Rodaway A., Takeda H., Koshida S., Broadbent J., Price B.M.J.,
RA Smith J.C., Patient R., Holder N.;
RT "Induction of the mesoderm in the zebrafish germ ring by yolk cell
RT derived TGF-(beta) family signals and discrimination of mesoderm and
RT endoderm by FGF.";
RL Development 126:3067-3077(1999).
DR EMBL; AJ242515; CAB43400.1; -.
DR HSSP; PI7679; IGPF.
DR ZFIN; ZDB-GENE-980526-340; gata5.
DR InterPro; IPR001164; hrip_like.
DR InterPro; IPR0006679; znf_GATA.
DR Pfam; PF00320; GATA; 2.
DR PRINTS; PR00619; GATAZFINGER.
DR SMART; SM00401; ZnF_GATA; 2.
DR PROSITE; PS00344; GATA_ZNFINGER_1; 2.
DR PROSITE; PS01114; GATA_ZNFINGER_2; 1.
SQ SEQUENCE 383 AA; 41479 MW; 7F752664185588AF CRC64;

Query Match 69.9%; Score 195; DB 13; Length 383;
Best Local Similarity 64.0%; Pred. No. 3.2e-18;
Matches 32; Conservative 10; Mismatches 8; Indels 0; Gaps

Qy 1 CTNCGKTTPLWRNPQGOPLCNACGLFLKLHGVRPLSLKTDVKKQR 50
|||||:||||| |:|:|||||: ||||| |:|:| | :| :| :|
Db 241 CTNCTSTTTTLWRRNAEGEPVCNAGLYMKLHGVNRPLAMRKESIQTKR 290
|||||:||||| |:|:|||||: ||||| |:|:~| :| :| :|

Search completed: January 3, 2003, 19:50:34
Job time : 49 secs

```


GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 3, 2003, 18:50:12 ; Search time 14 Seconds
(without alignments)
148.130 Million cell updates/sec

Title: US-09-725-010-3
Perfect score: 281
Sequence: 1 CTNCFQTPTPLRRNPDCQP.....LHGVRPLSLKTDVKKRNR 50

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Swissprot_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	281	100.0	1036	1	NIT2_NEUCR
2	277	98.6	725	1	AREA_PENCH
3	277	98.6	860	1	AREA_PENRO
4	277	98.6	865	1	NREA_PENRO
5	277	98.6	866	1	AREA_ASPOR
6	277	98.6	876	1	AREA_EMENI
7	277	98.6	882	1	AREA_ASPNG
8	277	98.6	971	1	AREA_GIBFU
9	269	95.7	956	1	NUT1_MAGGR
10	252	89.7	855	1	GAF1_SCHPO
11	231	82.2	510	1	GAT1_YEAST
12	211	75.1	426	1	ASB4_NEUCR
13	210	74.7	551	1	GZF3_YEAST
14	204	72.6	269	1	DA80_YEAST
15	201	71.5	730	1	GLN3_YEAST
16	195	69.4	388	1	GASB_XENLA
17	195	69.4	390	1	GASA_XENLA
18	193	68.7	540	1	PNR_DROME
19	190	67.6	404	1	GAT5_MOUSE
20	190	67.6	779	1	SRP_DROME
21	189	67.3	391	1	GAGA_XENLA
22	189	67.3	391	1	GAT5_CHICK
23	189	67.3	397	1	GAT5_HUMAN
24	188	66.9	387	1	GAT6_CHICK
25	188	66.9	391	1	GAT6_XENLA
26	188	66.9	392	1	GAT4_XENLA
27	188	66.9	441	1	GAT6_RAT
28	188	66.9	444	1	GAT6_MOUSE
29	188	66.9	449	1	GAT6_HUMAN
30	187	66.5	950	1	URB1_USTMA
31	185	65.8	380	1	GAT4_CHICK
32	183	65.1	413	1	GAT1_MOUSE
33	183	65.1	440	1	GAT4_MOUSE

34	183	65.1	440	1	GAT4_RAT
35	183	65.1	442	1	GAT4_HUMAN
36	182	64.8	359	1	GATA_XENLA
37	182	64.8	413	1	GAT1_HUMAN
38	182	64.8	416	1	GAT1_RAT
39	182	64.8	416	1	ELT1_CAPEL
40	181	64.4	364	1	GAT1B_XENLA
41	181	64.4	564	1	GAF2_SCHPO
42	180	64.1	452	1	GAT2_XENLA
43	180	64.1	486	1	GATC_DROME
44	179	63.7	304	1	GAT1_CHICK
45	179	63.7	466	1	GAT2_CHICK

ALIGNMENTS

RESULT 1

NIT2_NEUCR					
ID	NIT2_NEUCR	STANDARD;	PRT;	1036 AA.	
AC	P19212;				
DT	01-NOV-1990 (Rel. 16, Created)				
DT	01-NOV-1990 (Rel. 16, Last sequence update)				
DT	16-OCT-2001 (Rel. 40, Last annotation update)				
DE	Nitrogen catabolic enzyme regulatory protein.				
GN	NIT-2				
OS	Neurospora crassa.				
OC	Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;				
OX	Sordariales; Sordariaceae; Neurospora.				
NCBI_TaxID	5141;				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=74-OR23-1A;				
RA	MEDLINE=90158568; PubMed=2137552;				
RT	Fu Y.-H., Marzluf G.A.;				
RT	"nit-2, the major nitrogen regulatory gene of Neurospora crassa,"				
RT	encodes a protein with a putative zinc finger DNA-binding domain.";				
RL	Mol. Cell. Biol. 10:1056-1065(1990).				
FN	[2]				
RP	MUTAGENESIS.				
RA	MEDLINE=91186820; PubMed=2150539;				
RT	Fu Y.-H., Marzluf G.A.;				
RT	"Site-directed mutagenesis of the 'zinc finger' DNA-binding domain of				
RT	the nitrogen-regulatory protein NIT2 of Neurospora.";				
RL	Mol. Microbiol. 4:1847-1852(1990).				
CC	FUNCTION: MAJOR NITROGEN REGULATORY PROTEIN. DURING CONDITIONS				
CC	OF NITROGEN LIMITATION IT TURNS ON THE EXPRESSION OF GENES FOR				
CC	ENZYMES WHICH ARE REQUIRED FOR THE USE OF A VARIETY OF SECONDARY				
CC	NITROGEN SOURCES, INCLUDING NITRATES, PURINES, AMINO ACIDS, AND				
CC	PROTEINS.				
CC	!- SUBCELLULAR LOCATION: Nuclear.				
CC	!- INDUCTION: BY LACK OF A PRIMARY NITROGEN SOURCE.				
CC	!- SIMILARITY: CONTAINS 1 GATA-TYPE ZINC FINGER.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/				
CC	or send an email to license@isb-sib.ch).				
CC	-----				
CC	EMBL; M33956; AAB03891.1; -				
CC	PIR; A34755; A34755.				
CC	HSSP; P17429; AGAT.				
CC	TRANSFAC; T00627; -				
CC	InterPro; IPR000679; Znf_GATA.				
CC	Pfam; PF00320; GATA; 1.				
CC	PRINTS; PR00619; GATAZNFINGER.				
CC	SMART; SM00401; Znf_GATA; 1.				
CC	PROSITE; PS00344; GATA_ZN_FINGER_1; 1.				
CC	PROSITE; PS50114; GATA_ZN_FINGER_2; 1.				
CC	Transcription regulation; Activator; DNA-binding; Zinc-finger;				
KW					

P46152	rattus norv
P43694	homo sapien
P23767	xenopus lae
P15976	homo sapien
P43429	rattus norv
P28515	caenorhabdl
P23768	xenopus lae
Q10134	schizosacch
P23770	xenopus lae
P91623	drosophila
P18778	gallus gall
P23824	gallus gall

KW Nuclear protein; Nitrate assimilation; Repeat.
FT DOMAIN 49 110 3 X APPROXIMATE REPEATS.
FT REPEAT 49 55 1.
FT REPEAT 87 92 2.
FT REPEAT 105 110 3.
FT ZN_FING 743 767 GATA-TYPE.
FT MUTAGEN 743 746 CTNC->STNG: ABOLISHES DNA-BINDING.
FT MUTAGEN 755 756 RR->GG: ABOLISHES DNA-BINDING.
FT MUTAGEN 765 766 NA->DV: ABOLISHES DNA-BINDING.
FT MUTAGEN 768 769 GL->DV: ABOLISHES DNA-BINDING.
FT MUTAGEN 789 790 KR->NS: ABOLISHES DNA-BINDING.
SQ SEQUENCE 1036 AA; 109310 MW; 5FE4992B1C223514 CRC64;

Query Match 100.0%; Score 281; DB 1; Length 1036;
Best Local Similarity 100.0%; Pred. No. 2.4e-28;
Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTNCFCTTTLWRRNPDGQPLCNACGLFLKLGHVVRPLSLKTDVIKKNR 50
|||||
DB 743 CTNCFCTTTLWRRNPDGQPLCNACGLFLKLGHVVRPLSLKTDVIKKNR 792

RESULT 2
AREA_PENCH
ID AREA_PENCH STANDARD; PRT; 725 AA.
AC Q01582;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Nitrogen regulatory protein area (Nitrogen regulator nre).
GN AREA OR NRE.
OS Penicillium chrysogenum.

OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eukotiales; Trichocomaceae; mitosporic Trichocomaceae; Penicillium.
OX NCBI_TaxID=5076;
RN [1]
RP SEQUENCE FROM N.A.

RA Haas H., Bauer B., Redl B., Stoeffler G., Marzluf G.A.;
RT "Molecular cloning and analysis of nre, the major nitrogen regulatory gene of Penicillium chrysogenum.";
RL Curr. Genet. 27:150-158(1995).
CC -!- FUNCTION: MAJOR NITROGEN REGULATORY PROTEIN.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- SIMILARITY: CONTAINS 1 GATA-TYPE ZINC FINGER.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to license@isb-sib.ch).

DR EMBL; U02612; AAA83400.1; -.
DR HSSP; P17429; 4GAT.
DR InterPro; IPR000679; Znf_GATA.
DR Pfam; PF00320; GATA; 1.
DR PRINTS; PRO0619; GATAZNFINGER.
DR SMART; SM00401; Znf_GATA; 1.
DR PROSITE; PS00344; GATA_2N_FINGER_1; 1.
DR PROSITE; PS0114; GATA_2N_FINGER_2; 1.
KW Transcription regulation; Activator; DNA-binding; Zinc-finger;
KW Nuclear protein; Nitrate assimilation.
FT ZN_FING 525 549 GATA-TYPE.
SQ SEQUENCE 725 AA; 76848 MW; 45C24148F81F8D43 CRC64;

Query Match 98.6%; Score 277; DB 1; Length 725;
Best Local Similarity 98.0%; Pred. No. 5.7e-28;
Matches 49; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTNCFCTTTLWRRNPDGQPLCNACGLFLKLGHVVRPLSLKTDVIKKNR 50
|||||

DB 525 CTNCFCTTTLWRRNPDGQPLCNACGLFLKLGHVVRPLSLKTDVIKKNR 574

RESULT 3

AREA_PENRO
ID AREA_PENRO STANDARD; PRT; 860 AA.
AC Q13508;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Nitrogen regulatory protein area (Nitrogen regulator nmc).
GN AREA OR NMC.
OS Penicillium roqueforti.

OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eukotiales; Trichocomaceae; mitosporic Trichocomaceae; Penicillium.
OX NCBI_TaxID=5082;
RN [1]
RP SEQUENCE FROM N.A.

RC STRAIN=P2;

RA Gente S., Poussereau N., Fevre M.;

RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.

CC -!- FUNCTION: MAJOR NITROGEN REGULATORY PROTEIN.

CC -!- SUBCELLULAR LOCATION: Nuclear.

CC -!- SIMILARITY: CONTAINS 1 GATA-TYPE ZINC FINGER.

CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to license@isb-sib.ch).

DR EMBL; AJ001530; CAA04815.1; -.
DR HSSP; P17429; 4GAT.
DR InterPro; IPR000679; Znf_GATA.
DR Pfam; PF00320; GATA; 1.
DR PRINTS; PRO0619; GATAZNFINGER.

DR SMART; SM00401; Znf_GATA; 1.

DR PROSITE; PS00344; GATA_2N_FINGER_1; 1.

DR PROSITE; PS0114; GATA_2N_FINGER_2; 1.

KW Transcription regulation; Activator; DNA-binding; Zinc-finger;

KW Nuclear protein; Nitrate assimilation.

FT ZN_FING 660 684 GATA-TYPE.

SQ SEQUENCE 860 AA; 91615 MW; 2B8BF0DDC1BA68 CRC64;

Query Match 98.6%; Score 277; DB 1; Length 860;
Best Local Similarity 98.0%; Pred. No. 6.7e-28;
Matches 49; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTNCFCTTTLWRRNPDGQPLCNACGLFLKLGHVVRPLSLKTDVIKKNR 50
|||||

DB 660 CTNCFCTTTLWRRNPDGQPLCNACGLFLKLGHVVRPLSLKTDVIKKNR 709

RESULT 4

NRFA_PENUR

ID NRFA_PENUR STANDARD; PRT; 865 AA.

AC Q92269;

DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Nitrogen regulatory protein NRFA.

GN NRFA.

OS Penicillium urticae.

OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;

OC Eukotiales; Trichocomaceae; mitosporic Trichocomaceae; Penicillium.

OX NCBI_TaxID=29844;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=NRRL 2159A;

RA Ellis C.M.;

RL Thesis (1996), University of Calgary, Canada.

```
CC -!- FUNCTION: MAJOR NITROGEN REGULATORY PROTEIN.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- SIMILARITY: CONTAINS 1 GATA-TYPE ZINC FINGER.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U53137; AAB17740.1; -.
DR HSSP; P17429; 4GAT.
DR TRANSFAC; T02828; -.
DR InterPro; IPR000679; Znf_GATA.
DR Pfam; PF00320; GATA; 1.
DR PRINTS; PR00619; GATAZNFINGER.
DR SMART; SM00401; ZNF_GATA; 1.
DR PROSITE; PS00344; GATA_ZN_FINGER_1; 1.
DR PROSITE; PS00114; GATA_ZN_FINGER_2; 1.
DR PROSITE; PS00114; GATA_ZN_FINGER_2; 1.
KW Transcription regulation; Activator; DNA-binding; Zinc-finger;
KW Nuclear protein; Nitrate assimilation.
FT ZN_FING 665 GATA-TYPE.
SQ SEQUENCE 865 AA; 92407 MW; CC901BED187D8EEF CRC64;

Query Match 98.6%; Score 277; DB 1; Length 865;
Best Local Similarity 98.0%; Pred. No. 6.7e-28;
Matches 49; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTNCFQTPTPLWRNPDCGQPLCNACGLFLKLGWVRPLSLKTDVTKRNR 50
DB 665 CTNCFQTPTPLWRNPDCGQPLCNACGLFLKLGWVRPLSLKTDVTKRNR 714

RESULT 5
AREA_ASPO
ID AREA_ASPO STANDARD; PRT; 866 AA.
AC O13415;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Nitrogen regulatory protein area.
GN AREA.
OS Aspergillus oryzae.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
OX NCBI_TaxID=5062;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-IFO 4177;
RA Christensen T., Hynes M.J., Davis M.A.;
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: MAJOR NITROGEN REGULATORY PROTEIN. POSITIVELY ACTING
CC REGULATORY GENE OF NITROGEN METABOLITE REPRESSION (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- SIMILARITY: CONTAINS 1 GATA-TYPE ZINC FINGER.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AJ002968; CAA05776.1; -.
DR HSSP; P17429; 4GAT.
DR InterPro; IPR000679; Znf_GATA.
DR Pfam; PF00320; GATA; 1.
DR PRINTS; PR00619; GATAZNFINGER.
DR SMART; SM00401; ZNF_GATA; 1.
DR PROSITE; PS00344; GATA_ZN_FINGER_1; 1.

DR PROSITE; PS00114; GATA_ZN_FINGER_2; 1.
KW Transcription regulation; Activator; DNA-binding; Zinc-finger;
KW Nuclear protein; Nitrate assimilation.
FT ZN_FING 668 GATA-TYPE.
SQ SEQUENCE 866 AA; 92882 MW; C09A03EB12E3FBB4 CRC64;

Query Match 98.6%; Score 277; DB 1; Length 866;
Best Local Similarity 98.0%; Pred. No. 6.7e-28;
Matches 49; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTNCFQTPTPLWRNPDCGQPLCNACGLFLKLGWVRPLSLKTDVTKRNR 50
DB 664 CTNCFQTPTPLWRNPDCGQPLCNACGLFLKLGWVRPLSLKTDVTKRNR 713

RESULT 6
AREA_EMENI
ID AREA_EMENI STANDARD; PRT; 876 AA.
AC P17429;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Nitrogen regulatory protein area.
GN AREA.
OS Emericella nidulans (Aspergillus nidulans).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; Emericella.
OX NCBI_TaxID=5072;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90228331; PubMed=1970293;
RA Kudla B., Caddick M.X., Langdon T., Martinez-Rossi N.M.,
RA Bennett C.F., Sibley S., Davies R.W., Arst H.N. Jr.;
RT "The regulatory gene area mediating nitrogen metabolite repression in
RT Aspergillus nidulans. Mutations affecting specificity of gene
RT activation alter a loop residue of a putative zinc finger.";
RL EMBO J. 9:1355-1364(1990).
RN [2]
RP REVISIONS.
RX MEDLINE=96123430; PubMed=8596437;
RA Langdon T., Seerins A., Ravagnani A., Caddick M.X., Arst H.N. Jr.;
RT "Mutational analysis reveals dispensability of the N-terminal region
RT of the Aspergillus transcription factor mediating nitrogen metabolite
RT repression.";
RL Mol. Microbiol. 17:877-888(1995).
RN [3]
RP STRUCTURE BY NMR OF 662-727.
RX MEDLINE=98202574; PubMed=9533883;
RA Starich M.R., Wikstrom M., Arst H.N. Jr., Clore G.M.,
RA Gronenborn A.M.;
RT "The solution structure of a fungal AREA protein-DNA complex: an
RT alternative binding mode for the basic carboxyl tail of GATA
RT factors.";
RL J. Mol. Biol. 277:605-620(1998).
CC -!- FUNCTION: TRANSCRIPTION ACTIVATOR THAT MEDIATES NITROGEN
CC METABOLITE REPRESSION IN A. NIDULANS. ACTIVATE THE TRANSCRIPTION
CC OF UAPA.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- SIMILARITY: CONTAINS 1 GATA-TYPE ZINC FINGER.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X52491; CAA36731.1; -.
DR PIR; S10017; S10017.
DR PDB; 4GAT; 28-JAN-98.
DR PDB; 5GAT; 28-JAN-98.
DR PDB; 6GAT; 28-JAN-98.
```

```
DR PDB: 7GAT; 28-JAN-98.
DR TRANSFAC; T02533; -.
DR InterPro; IPR000679; Znf_GATA.
DR Pfam; PF00320; GATA; 1.
DR PRINTS; PR00619; GATAZNFINGER.
DR SMART; SM00401; ZNF_GATA; 1.
DR PROSITE; PS00344; GATA_ZN_FINGER_1; 1.
DR PROSITE; PS00344; GATA_ZN_FINGER_2; 1.
KW Transcription regulation; Activator; DNA-binding; Zinc-finger;
KW Nuclear protein; Nitrate assimilation; 3D-structure.
FT ZN_FING 673 697 GATA-TYPE.
FT DNA_BIND 721 742 H-T-H MOTIF (PROBABLE).
SQ SEQUENCE 876 AA; 94195 MW; 9ADC2273EE536F98 CRC64;

Query Match 98.6%; Score 277; DB 1; Length 876;
Best Local Similarity 98.0%; Pred. No. 6.8e-28;
Matches 49; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTNCFQTTPPLWRNPDPGQPLCNACGLFLKLGHVVRPLSLKTDVIKKRNR 50
DB 673 CTNCFQTTPPLWRNPDPGQPLCNACGLFLKLGHVVRPLSLKTDVIKKRNR 722

RESULT 7
AREA ASPNG
ID AREA ASPNG STANDARD; PRT; 882 AA.
AC O13412;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Nitrogen regulatory protein area.
GN AREA.
OS Aspergillus niger.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eutrotales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
OX NCBI_TaxID=5061;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=N402;
RA Macabe A.P., Vanhanen S.A.S., Sollewijn Gelpke M.,
RA van de Vondervoort P., Arst H.N., Visser J.;
RL Submitted (SEP-1994) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: MAJOR NITROGEN REGULATORY PROTEIN. POSITIVELY ACTING
CC REGULATORY GENE OF NITROGEN METABOLITE REPRESSION.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- SIMILARITY: CONTAINS 1 GATA-TYPE ZINC FINGER.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X81998; CAA57524.1; -.
CC HSP; P17429; 4GAT.
CC InterPro; IPR000679; Znf_GATA.
CC Pfam; PF00320; GATA; 1.
CC PRINTS; PR00619; GATAZNFINGER.
CC SMART; SM00401; ZNF_GATA; 1.
CC PROSITE; PS00344; GATA_ZN_FINGER_1; 1.
CC PROSITE; PS00344; GATA_ZN_FINGER_2; 1.
KW Transcription regulation; Activator; DNA-binding; Zinc-finger;
KW Nuclear protein; Nitrate assimilation.
FT ZN_FING 676 700 GATA-TYPE.
SQ SEQUENCE 882 AA; 94518 MW; 97A502936B94E5AE CRC64;

Query Match 98.6%; Score 277; DB 1; Length 882;
Best Local Similarity 98.0%; Pred. No. 6.8e-28;
Matches 49; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTNCFQTTPPLWRNPDPGQPLCNACGLFLKLGHVVRPLSLKTDVIKKRNR 50
```

```
DB 676 CTNCFQTTPPLWRNPDPGQPLCNACGLFLKLGHVVRPLSLKTDVIKKRNR 725

RESULT 8
AREA GIBFU
ID AREA GIBFU STANDARD; PRT; 971 AA.
AC P78688;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Nitrogen regulatory protein area.
GN AREA.
OS Gibberella fujikuroi (Bakanae and foot rot disease fungus) (Fusarium
OS moniliforme).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Hypocreales; Nectriaceae; Gibberella.
OX NCBI_TaxID=5127;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=m567;
RX MEDLINE=99168774; PubMed=10071216;
RA tudzynski B., Homann V., Feng B., Marzluf G.A.;
RT "Isolation, characterization and disruption of the area nitrogen
RT regulatory gene of Gibberella fujikuroi.";
RL Mol. Gen. Genet. 261:106-114(1999).
CC -!- FUNCTION: MAJOR NITROGEN REGULATORY PROTEIN. POSITIVELY ACTING
CC REGULATORY GENE OF NITROGEN METABOLITE REPRESSION (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- SIMILARITY: CONTAINS 1 GATA-TYPE ZINC FINGER.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Y11006; CAA71897.1; -.
CC HSP; P17429; 4GAT.
CC InterPro; IPR000679; Znf_GATA.
CC Pfam; PF00320; GATA; 1.
CC PRINTS; PR00619; GATAZNFINGER.
CC SMART; SM00401; ZNF_GATA; 1.
CC PROSITE; PS00344; GATA_ZN_FINGER_1; 1.
CC PROSITE; PS00344; GATA_ZN_FINGER_2; 1.
KW Transcription regulation; Activator; DNA-binding; Zinc-finger;
KW Nuclear protein; Nitrate assimilation.
FT ZN_FING 694 718 GATA-TYPE.
SQ SEQUENCE 971 AA; 103580 MW; 887DD882141C7453 CRC64;

Query Match 98.6%; Score 277; DB 1; Length 971;
Best Local Similarity 98.0%; Pred. No. 7.5e-28;
Matches 49; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTNCFQTTPPLWRNPDPGQPLCNACGLFLKLGHVVRPLSLKTDVIKKRNR 50
DB 694 CTNCFQTTPPLWRNPDPGQPLCNACGLFLKLGHVVRPLSLKTDVIKKRNR 743

RESULT 9
NUTL_MAGGR
ID NUTL_MAGGR STANDARD; PRT; 956 AA.
AC Q01168;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Nitrogen regulatory protein NUTL.
GN NUTL.
OS Magnaporthe grisea (Rice blast fungus) (Pyricularia grisea).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariomycetes incertae sedis; Magnaporthaceae; Magnaporthe.
```



```

zipper.";
RL Curr. Genet. 21:301-307(1992).
[4]
RR SEQUENCE FROM N.A.
RA Urrestazu L.A., Jauniaux J.-C.;
RA Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: NEGATIVE REGULATOR OF MULTIPLE NITROGEN CATABOLIC
CC GENES INCLUDING THE ALLANTOIN PATHWAY GENES.
CC
CC -!- SUBCELLULAR LOCATION: Nuclear
CC -!- INDUCTION: SENSITIVE TO NITROGEN CATABOLITE REPRESSION.
CC -!- SIMILARITY: CONTAINS 1 GATA-TYPE ZINC FINGER.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M7821; AAA34556.1; -
CC DR EMBL: X60199; CAA42757.1; -
CC DR EMBL: Z28259; CAA82107.1; -
CC DR EMBL: Z28258; CAA82106.1; -
CC DR PIR: S22781; S22781.
CC DR HSSP: P17429; 4GAT.
CC DR TRANSFAC: T02411; -.
CC DR SGD: S0001742; DAL80.
CC DR InterPro: IPR000679; znf_GATA.
CC DR Pfam: PF00320; GATA; 1.
CC DR PRINTS: PR00619; GATAZNFINGER.
CC DR SMART: SM00401; znf_GATA; 1.
CC DR PROSITE: PS00344; GATA_ZN_FINGER_1; 1.
CC DR PROSITE: PS01114; GATA_ZN_FINGER_2; 1.
CC KW Transcription regulation; Repressor; DNA-binding; zinc-finger;
CC KW Nuclear protein; Nitrate assimilation.
CC ZN_FING 31 55
CC FT DOMAIN 79 100 ASN-RICH.
CC FT DOMAIN 101 108 ARG/LYS-RICH (BASIC).
CC FT CONFLICT 6 6 S -> L (IN REF. 3).
CC FT CONFLICT 207 207 V -> I (IN REF. 3).
CC SEQUENCE 269 AA; 30166 MW; F37BEB10038599EC CRC64;
CC
CC Query Match 72.6%; Score 204; DB 1; Length 269;
CC Best Local Similarity 74.0%; Pred. No. 5.4e-19;
CC Matches 37; Conservative 2; Mismatches 11; Indels 0; Gaps 0;
CC
CC QY 1 CTNCFQTPTPLWRNPDGQPLCNACGLFLKLHGVVRPLSLKTDVIKKRNR 50
CC | ||||| ||||| |
CC DB 31 CQNCFTVKTPLWRDEHGTVLCNACGLFLKLHGEPRPISLKTDTIKSRNR 80
CC | ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
CC
CC RESULT 15
CC GLN3_YEAST
CC ID GLN3_YEAST STANDARD; PRT; 730 AA.
CC AC P18494;
CC DT 01-NOV-1990 (Rel. 16, Created)
CC DT 01-FEB-1995 (Rel. 31, Last sequence update)
CC DT 16-OCT-2001 (Rel. 40, Last annotation update)
CC DE Nitrogen regulatory protein GLN3.
CC GN GLN3 OR YER040W.
CC OS Saccharomyces cerevisiae (Baker's yeast).
CC OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
CC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
CC OX NCBI_TaxID=4932;
CC RN [1]
CC SEQUENCE FROM N.A.
CC RP MEDLINE=92049353; PubMed=1682800;
CC RA Minehart P.L., Magasanik B.;
CC "Sequence and expression of GLN3, a positive nitrogen regulatory gene
CC of Saccharomyces cerevisiae encoding a protein with a putative zinc
CC finger DNA-binding domain.";
CC Mol. Cell. Biol. 11:6216-6228(1991).
CC RL

```

```

RN SEQUENCE FROM N.A.
RC STRAIN-S288C / AB972;
RA Dietrich F.S., Mulligan J.T., Hennessey K.M., Allen E., Araujo R.,
RA Aviles E., Berno A., Brennan T., Carpenter J., Chen E., Cherry J.M.,
RA Chung E., Duncan M., Guzman E., Hartzell G., Hunnicke-Smith S.,
RA Hyman R., Kayser A., Komp C., Lashkari D., Lew H., Lin D.,
RA Mosedale D., Nakahara K., Namath A., Norgren R., Oefner P., Oh C.,
RA Petel F.X., Roberts D., Sehl P., Schramm S., Shogren T., Smith V.,
RA Taylor P., Wei Y., Yelton M., Botstein D., Davis R.W.;
RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: POSITIVE NITROGEN REGULATORY PROTEIN. REQUIRED FOR
CC THE ACTIVATION OF TRANSCRIPTION OF A NUMBER OF GENES (INCLUDING
CC THE ALLANTOIN PATHWAY GENES) IN RESPONSE TO THE REPLACEMENT OF
CC GLUTAMINE BY GLUTAMATE AS SOURCE OF NITROGEN. BINDS THE NITROGEN
CC UPSTREAM ACTIVATION SEQUENCE OF GLN1, THE GENE ENCODING GLUTAMINE
CC SYNTHETASE. URE2 MAY CATALYTICALLY INACTIVATE GLN3 IN RESPONSE TO
CC AN INCREASE IN THE INTRACELLULAR CONCENTRATION OF GLUTAMINE..
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- SIMILARITY: CONTAINS 1 GATA-TYPE ZINC FINGER.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M35267; AAA34645.1; -
DR EMBL; U18796; AAB64575.1; -
DR PIR; S22280; S22280.
DR HSSP; PI7429; 4GAT.
DR TRANSFAC; T02818; -.
DR SGD; S0000842; GLN3.
DR InterPro; IPR000679; Znf_GATA.
DR Pfam; PF00320; GATA.1.
DR PRINTS; PR00619; GATAZNFINGER.
DR SMART; SM00401; Znf_GATA.1.
DR PROSITE; PS00344; GATA_ZN_FINGER_1; 1.
DR PROSITE; PS01114; GATA_ZN_FINGER_2; 1.
KW Transcription regulation; Activator; DNA-binding; Zinc-finger;
KW Nuclear protein; Nitrate assimilation.
FT ZN_FING 306 330 GATA-TYPE.
FT DOMAIN 351 361 ARG/LYS-RICH (BASIC).
FT CONFLICT 474 474 P -> G (IN REF. 1).
SQ SEQUENCE 730 AA; 79382 MW; 3159E1844469942E CRC64;

Query Match
Best Local Similarity 71.5%; Score 201; DB 1; Length 730;
Matches 36; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

QY 1 CYNCFQTPTPLWRRNPDPGLNACGLFLKHLGVVVRPLSLKTDVIKKR 48
Db 306 CFNCFTKFTPLWRRSPGNTLCNACGLFQKHLGTMRLSLASDVIIKR 353

```

Search completed: January 3, 2003, 19:48:37
Job time : 14 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 3, 2003, 19:46:47 ; Search time 25 seconds
(without alignments)
192.269 Million cell updates/sec

Title: US-09-725-010-3

Perfect score: 281

Sequence: 1 CINCFTQTTPWRRNPDPG.....LHGVRPLSLKTDVIKRRN 50

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR_73:*

1: pir1:*

2: pir2:*

3: pir3:*

4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Length	ID	Description
1	281	100.0	1036	1 A34755	nitrogen regulator
2	277	98.6	862	2 S51493	major nitrogen reg
3	277	98.6	876	1 A57988	regulatory protein
4	252	89.7	855	2 T41336	probable nitrogen
5	231	82.2	510	2 S56233	probable membrane
6	210	74.7	551	2 S53377	GZF3 protein - yea
7	204	72.6	269	2 S22781	transcription fact
8	201	71.5	730	2 S50543	GLN3 protein - yea
9	195	69.4	388	2 I51420	transcription fact
10	195	69.4	390	2 I51419	transcription fact
11	190	67.6	779	1 S04382	box A-binding fact
12	189	67.3	391	2 I50702	transcription fact
13	188	66.9	387	2 I50703	transcription fact
14	188	66.9	391	2 A49645	transcription fact
15	187	66.5	950	2 S27473	transcription fact
16	185	65.8	380	2 I50701	URB1 protein - sm
17	183	65.1	413	2 S04655	transcription fact
18	183	65.1	439	1 A48099	transcription fact
19	183	65.1	439	2 I57561	transcription fact
20	183	65.1	440	2 I61183	transcription fact
21	182	64.8	359	2 A41602	transcription fact
22	182	64.8	413	1 A34888	transcription fact
23	182	64.8	413	2 S48756	transcription fact
24	182	64.8	416	1 A41267	transcription fact
25	182	64.8	454	2 T26296	hypothetical prote
26	181	64.4	364	2 B41602	transcription fact
27	181	64.4	564	2 T38291	GATA-type transcri
28	180	64.1	452	2 C41602	transcription fact
29	180	64.1	486	1 A57601	transcription fact

30 179 63.7 118 2 B48099 transcription fact
31 179 63.7 304 2 A32993 transcription fact
32 179 63.7 466 2 A36389 transcription fact
33 179 63.7 474 2 A41782 transcription fact
34 179 63.7 480 2 A40815 GATA-transcription
35 179 63.7 532 2 JG6170 transcription fact
36 178 63.3 435 2 D41602 transcription fact
37 178 63.3 443 2 B39794 transcription fact
38 178 63.3 444 1 A39794 transcription fact
39 178 63.3 444 2 B36389 transcription fact
40 175 62.3 564 2 T43298 transcription fact
41 171 60.9 241 2 S53812 BmGATA beta isofo
42 171 60.9 327 2 S53811 transcription fact
43 171 60.9 509 2 A53741 transcription fact
44 150.5 53.6 433 2 A56953 transcription fact
45 150.5 53.6 613 2 T19677 hypothetical prote

ALIGNMENTS

RESULT 1

A34755

nitrogen regulatory protein nit-2 - Neurospora crassa

C:Species: Neurospora crassa

C>Date: 13-Jul-1990 #sequence_revision 26-Jul-1996 #text_change 16-Feb-1997

C:Accession: A34755

R:Fu, Y.H.; Marzluf, G.A.

Mol. Cell. Biol. 10, 1056-1065, 1990

A>Title: nit-2, the major nitrogen regulatory gene of Neurospora crassa, encodes a pr

A:Reference number: A34755; MUID:90158568; PMID:2137552

A:Accession: A34755

A:Molecule type: DNA; mRNA

A:Residues: 1-1036 <FU>

A:Cross-references: GB:M33956

C:Genetics:

A:Introns: 209/2; 335/3

C:Superfamily: nitrogen regulatory protein nit-2; GATA-type zinc finger homology

C:Keywords: DNA binding; transcription regulation; zinc finger

F:740-793/Domain: GATA-type zinc finger homology <GZF>

F:743-767/Region: zinc finger GATA motif

Query Match 100.0% Score 281; DB 1; Length 1036;
Best Local Similarity 100.0%; Pred. No. 1e-27;
Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CINCFTQTTPWRRNPDPGQPLCNACGLFLKLGVRPLSLKTDVIKRRN 50

|||||

Db 743 CINCFTQTTPWRRNPDPGQPLCNACGLFLKLGVRPLSLKTDVIKRRN 792

RESULT 2

S51493

major nitrogen regulation protein - Penicillium chrysogenum (strain Q176)

C:Species: Penicillium chrysogenum

A:Variety: strain Q176

C>Date: 28-Oct-1996 #sequence_revision 13-Mar-1997 #text_change 08-Dec-2000

C:Accession: S51493

R:Haas, H.; Bauer, B.; Redl, B.; Stoeffler, G.; Marzluf, G.A.

Curr. Genet. 27, 150-158, 1995

A>Title: Molecular cloning and analysis of nre, the major nitrogen regulatory gene of

A:Reference number: S51493; MUID:95308537; PMID:7798718

A:Accession: S51493

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-862 <HAA>

A:Cross-references: EMBL:U02612

C:Genetics:

A:Introns: 123/2

C:Superfamily: nitrogen regulatory protein nit-2; GATA-type zinc finger homology

C:Keywords: zinc finger

F:659-712/Domain: GATA-type zinc finger homology <GZF>

Query Match 98.6%; Score 277; DB 2; Length 862;
Best Local Similarity 98.0%; Pred. No. 2.8e-27;
Matches 49; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTNCFQTTPPLWRNPDGQPLCNACGLFLKLGHVVRPLSLKTDVIKKNR 50
|||||
DB 662 CTNCFQTTPPLWRNPDGQPLCNACGLFLKLGHVVRPLSLKTDVIKKNR 711
|||||

RESULT 3
A57988
regulatory protein areA - Emericella nidulans
C:Species: Emericella nidulans, Aspergillus nidulans
C:Date: 26-Jul-1996 #sequence_revision 26-Jul-1996 #text_change 16-Jun-2000
C:Accession: A57988; S10017; S70168; S72883
R:Kudla, B.; Caddick, M.X.; Langdon, T.; Martinez-Rossi, N.M.; Bennett, C.F.; Sibley, S.
EMBO J. 9, 1355-1364, 1990
A:Title: The regulatory gene areA mediating nitrogen metabolite repression in Aspergillus
A:Reference number: S10017; MUID:90228331; PMID:1970293
A:Accession: A57988
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-876 <KUD>
A:Cross-references: EMBL:X52491; NID:g1019911; PIDN:CRA36731.1; PID:g1154625
A:Note: this sequence represents reinterpretation to include two exons
A:Accession: S10017
A:Molecule type: DNA
A:Residues: 158-876 <KU2>
A:Cross-references: EMBL:X52491
A:Note: this sequence represents the authors' original translation
R:Langdon, T.; Sheerins, A.; Ravagnani, A.; Gielkens, M.; Caddick, M.X.; Arst Jr., H.N.
Mol. Microbiol. 17, 877-888, 1995
A:Title: Mutational analysis reveals dispensability of the N-terminal region of the Aspe
A:Reference number: S70167; MUID:96123430; PMID:8556437
A:Accession: S70168
A:Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1-791, 'A', 793, 'T', 795, 'SPGTNS', 802-876 <LAN>
A:Cross-references: EMBL:X52491
R:Caddick, M.X.
submitted to the EMBL Data Library, October 1995
A:Reference number: S72883
A:Accession: S72883
A:Molecule type: DNA
A:Residues: 1-876 <CAD>
A:Cross-references: EMBL:X52491; NID:g1019911; PIDN:CRA36731.1; PID:g1154625
C:Genetics:
A:Gene: areA
A:Introns: 147/2
C:Function:
A:Description: mediates nitrogen metabolite repression
C:Superfamily: nitrogen regulatory protein nit-2; GATA-type zinc finger homology
C:Keywords: DNA binding; transcription regulation; zinc finger
F:670-723/Domain: GATA-type zinc finger homology <GZF>
F:673-697/Region: zinc finger GATA motif

Query Match 98.6%; Score 277; DB 1; Length 876;
Best Local Similarity 98.0%; Pred. No. 2.9e-27;
Matches 49; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTNCFQTTPPLWRNPDGQPLCNACGLFLKLGHVVRPLSLKTDVIKKNR 50
|||||
DB 673 CTNCFQTTPPLWRNPDGQPLCNACGLFLKLGHVVRPLSLKTDVIKKNR 722
|||||

RESULT 4
T41336
probable nitrogen regulatory transcription factor gaf1 - fission yeast (Schizosaccharomyces
C:Species: Schizosaccharomyces pombe
C:Date: 20-Oct-2000 #sequence_revision 08-Dec-2000 #text_change 08-Dec-2000
A:Accession: T41336; T41208; T43297
R:Lyne, M.; Rajandream, M.A.; Barrell, B.G.; Voicckaert, G.
submitted to the EMBL Data Library, January 1999

A:Reference number: Z21970
A:Accession: T41336
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-170 <LYN>
A:Cross-references: EMBL:AL035076; NID:g4107277; PIDN:CAA22647.1; PID:g4107278; GSPDB
A:Experimental source: strain 972h-; cosmid c417
R:Seeger, K.; Harris, D.; Wood, V.; Rajandream, M.A.; Barrell, B.G.
submitted to the EMBL Data Library, March 1999
A:Reference number: Z21928
A:Accession: T41208
A:Molecule type: DNA
A:Residues: 129-855 <SEE>
A:Cross-references: EMBL:AL049521; NID:g6318249; PIDN:CAB40003.1; PID:g4539591; GSPDB
R:Hoe, K.L.; Won, M.S.; Chung, K.S.; Park, S.K.; Kim, D.U.; Jang, Y.J.; Yoo, O.J.; Yo
Gene 215, 319-328, 1998
A:Title: Molecular cloning of gaf1, a Schizosaccharomyces pombe GATA factor, which ca
A:Reference number: Z22401; MUID:98382525; PMID:9714831
A:Accession: T43297
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 566-855 <HOE>
A:Cross-references: EMBL:L31601; NID:g710606; PIDN:AAC35593.1; PID:g710607
C:Genetics:
A:Gene: SPDB:SPCC417.01c; SPDB:SPCC1902.01; gaf1
A:Map position: 3R
C:Superfamily: GATA-type zinc finger homology
C:Keywords: transcription factor; transcription regulation; zinc finger
F:632-685/Domain: GATA-type zinc finger homology <GZF>

Query Match 89.7%; Score 252; DB 2; Length 855;
Best Local Similarity 88.0%; Pred. No. 4.4e-24;
Matches 44; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTNCFQTTPPLWRNPDGQPLCNACGLFLKLGHVVRPLSLKTDVIKKNR 50
|||||
DB 635 CTNCFQTTPPLWRNPDGQPLCNACGLFLKLGHVVRPLSLKTDVIKKNR 684
|||||

RESULT 5
S56233
probable membrane protein YFL021w - yeast (Saccharomyces cerevisiae)
C:Species: Saccharomyces cerevisiae
C:Date: 02-Sep-1995 #sequence_revision 19-Oct-1995 #text_change 06-Feb-1998
C:Accession: S56233
R:Murakami, Y.; Naitou, M.; Hagiwara, H.; Shibata, T.; Ozawa, M.; Sasanuma, S.I.; Sas
submitted to the EMBL Data Library, May 1995
A:Description: Analysis of the nucleotide sequence of chromosome VI from Saccharomyces
A:Reference number: S56186
A:Molecule type: DNA
A:Accession: S56233
A:Residues: 1-510 <MUR>
A:Cross-references: EMBL:D50617; NID:g836685; PID:d1009858; PID:g836733; MIPS:YFL021w
C:Genetics:
A:Gene: SGP:GAT1
A:Cross-references: SGD:S0001873; MIPS:YFL021w
A:Map position: 6L
C:Superfamily: GATA-type zinc finger homology
C:Keywords: transmembrane protein; zinc finger
F:14-30/Domain: transmembrane #status predicted <TM1>
F:307-360/Domain: GATA-type zinc finger homology <GZF>

Query Match 82.2%; Score 231; DB 2; Length 510;
Best Local Similarity 80.0%; Pred. No. 1.3e-21;
Matches 40; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY 1 CTNCFQTTPPLWRNPDGQPLCNACGLFLKLGHVVRPLSLKTDVIKKNR 50
|||||
DB 310 CSNCTTSTPLWRKDPKGLPLCNACGLFLKLGHVTRPLSLKTDIIKRRQ 359
|||||

RESULT 6
S53377

```

A:Molecule type: DNA
A:Residues: 1-473,'G',475-730 <MIN>
C:Genetics:
A:Gene: SGD:GLN3
A:Cross-references: SGD:S0000842; MIPS:YER040w
A:Map position: 5R
C:Superfamily: GATA-type zinc finger homology
C:Keywords: zinc finger
F:303-356/Domain: GATA-type zinc finger homology <GZF>
F:306-330/Region: zinc finger GATA motif

Query Match 71.5% Score 201; DB 2; Length 730;
Best Local Similarity 75.0%; Pred. No. 1.2e-17;
Matches 36; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

```

Qy 1 CTNCFQTTPWRRNPDGQPLCNACGLFLKLGHVVRPLSLKTVDIKR 48
 | | | | | : | | | | | : | | | | |
Dd 306 CFNCCKTEKTPWRRSPGNTICNACGLFOKLHGTMRLPSLSKSDVIKR 353

RESULT 9
I51420
transcription factor xGATA-4b - African clawed frog
C:Species: xenopus laevis (African clawed frog)
C:date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 08-Dec-2000
C:Accession: I51420
R:Kelley, C.; Blumberg, H.; Zon, L.I.; Evans, T.
Development 118, 817-827, 1993
A:title: GATA-4 is a novel transcription factor expressed in endocardium of the developing

RESULT 10
I51419
transcription factor xGATA-4a - African clawed frog
C:Species: Xenopus laevis (African clawed frog)
C:Date: 13-Sep-1996 #sequence_revision# 13-Sep-1996 #text_change 08-Dec-2000
C:Accession: I51419
R:Kelley, C.; Blumberg, H.; Zon, L.I.; Evans, T.
Development 118, 817-827, 1993
A:Ritte: GATA-4 is a novel transcription factor expressed in endocardium of the developing
A:Reference number: I51419; PMID:94357077; PMID:8076520

RESULT 11
S40382
box A-binding factor -, fruit fly (*Drosophila melanogaster*)
N:Alternate names: ABF; transcription factor GATAb
C:Species: *Drosophila melanogaster*
C>Date: 13-Jan-1995 #sequence_revision 06-Sep-1996 #text_change 16-Jul-1999
C:Accession: S40382
R:Abel, T.; Michelson, A.M.; Maniatis, T.
Development 119, 623-633, 1993
A:Title: A *Drosophila* GATA family member that binds to Adh regulatory sequences is expressed

A;Reference number: S40382; MUID:94244465; PMID:8187633
A;Accession: S40382
A;Molecule type: mRNA
A;Residues: 1-779 <ABE>
A;Cross-references: EMBL:X76217; NID:g441491; PIDN:CRA53807.1; PID:g441492
C;Comment: This transcriptional activator is the earliest known marker of the develop
C;Genetics:
A;Gene: FlyBase:srp
A;Cross-references: FlyBase:FBgn0003507
C;Superfamily: box A-binding factor; GATA-type zinc finger homology
C;Keywords: DNA binding; nucleus; transcription factor; transcription regulation; zin
F;316-369/Domain: GATA-type zinc finger homology <GZF>
F;319-343/Region: zinc finger GATA motif

RESULT 12

I50702

transcription factor GATA-5 - chicken

C/Species: Gallus gallus (chicken)

C/Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 26-Aug-1999

C/Accession: I50702

C/Laverriere, A.C.; MacNeill, C.; Mueller, C.; Poelmann, R.E.; Burch, J.B.; Evans, T. J. Biol. Chem. 269, 23177-23184, 1994

A/Title: GATA-4/5/6, a subfamily of three transcription factors transcribed in development

A/Reference number: A54720; MUID:94365018; PMID:8083222

A/Accession: I50702

A/Status: translated from GB/EMBL/DBDJ

A/Molecule type: mRNA

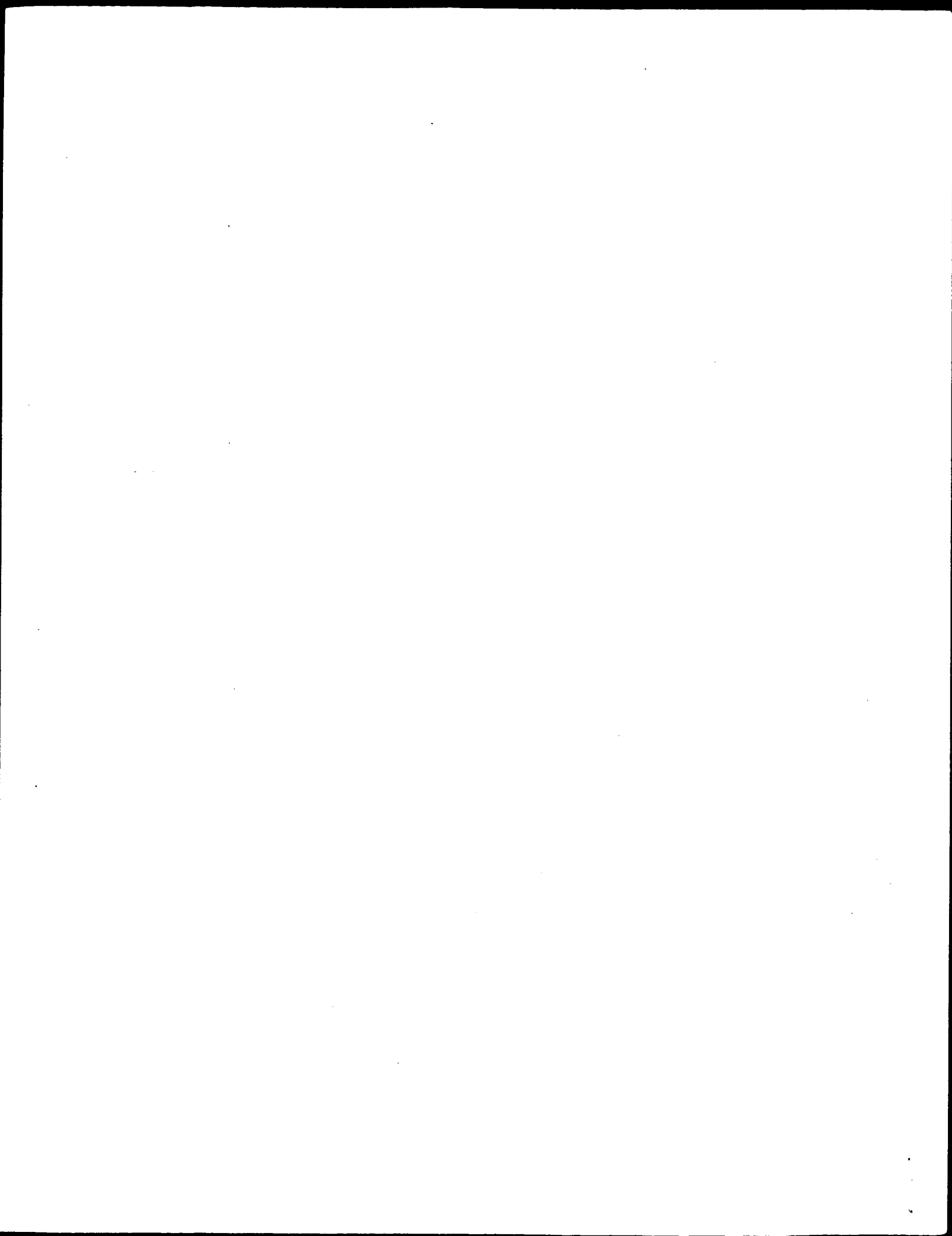
A/Residues: 1-391 <LAV>

A/Cross-references: EMBL:U1188; NID:g511481; PIDN:AAA57504.1; PID:g511482

C/Comment: This transcriptional activator is named for the core, GATA, of the nucleosome

C/Genetics:

RESULT 13
I50703
transcription factor GATA-6 - chicken
C:Species: Gallus gallus (chicken)
C:Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 26-Aug-1999
C:Accession: I50703
R:Lavertiere, A.C.; MacNeill, C.; Mueller, C.; Poelmann, R.E.; Burch, J.B.; Evans, T.
J. Biol. Chem. 269, 23177-23184, 1994
A:Title: GATA-4/5/6, a subfamily of three transcription factors transcribed in develop-
A:Reference number: A54720; MUID:94365018; PMID:8083222
A:Accession: I50703
A:Status: translated from GB/EMBL/DBDJ
A:Molecule type: mRNA
A:Residues: 1-387 <NAV>
A:Cross-references: EMBL:U11889; NID:g511483; PIDN:AAA57505.1; PID:g511484
C:Comment: This transcriptional activator is named for the core, GATA, of the nucleot-



GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 3, 2003, 19:48:43 ; Search time 187.5 Seconds
(without alignments)
171.929 Million cell updates/sec

Title: US-09-725-010-3
Perfect score: 281
Sequence: 1 CTNCFQTPTPLWRRNPDPQ.....LHGVRPLSLKTDVTKKRR 50

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 4569144 seqs, 644733110 residues

Total number of hits satisfying chosen parameters: 4569144

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Pending_Patents_AA_Main:*
1: /cgn2_6/ptodata/1/paa/PCTUS_COMB.pep.*
2: /cgn2_6/ptodata/1/paa/US06_COMB.pep.*
3: /cgn2_6/ptodata/1/paa/US07_COMB.pep.*
4: /cgn2_6/ptodata/1/paa/US080_COMB.pep.*
5: /cgn2_6/ptodata/1/paa/US081_COMB.pep.*
6: /cgn2_6/ptodata/1/paa/US082_COMB.pep.*
7: /cgn2_6/ptodata/1/paa/US083_COMB.pep.*
8: /cgn2_6/ptodata/1/paa/US084_COMB.pep.*
9: /cgn2_6/ptodata/1/paa/US085_COMB.pep.*
10: /cgn2_6/ptodata/1/paa/US086_COMB.pep.*
11: /cgn2_6/ptodata/1/paa/US087_COMB.pep.*
12: /cgn2_6/ptodata/1/paa/US088_COMB.pep.*
13: /cgn2_6/ptodata/1/paa/US089_COMB.pep.*
14: /cgn2_6/ptodata/1/paa/US090_COMB.pep.*
15: /cgn2_6/ptodata/1/paa/US091_COMB.pep.*
16: /cgn2_6/ptodata/1/paa/US092_COMB.pep.*
17: /cgn2_6/ptodata/1/paa/US093_COMB.pep.*
18: /cgn2_6/ptodata/1/paa/US094_COMB.pep.*
19: /cgn2_6/ptodata/1/paa/US095_COMB.pep.*
20: /cgn2_6/ptodata/1/paa/US096_COMB.pep.*
21: /cgn2_6/ptodata/1/paa/US097_COMB.pep.*
22: /cgn2_6/ptodata/1/paa/US098_COMB.pep.*
23: /cgn2_6/ptodata/1/paa/US099_COMB.pep.*
24: /cgn2_6/ptodata/1/paa/US100_COMB.pep.*
25: /cgn2_6/ptodata/1/paa/US101_COMB.pep.*
26: /cgn2_6/ptodata/1/paa/US102_COMB.pep.*
27: /cgn2_6/ptodata/1/paa/US60_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	281	100.0	1036	21	US-09-791-537-26785
2	281	100.0	1036	21	US-09-791-537-121036
3	277	98.6	66	21	US-09-791-537-44327
4	277	98.6	725	21	US-09-791-537-118817
5	277	98.6	860	21	US-09-791-537-125732
6	277	98.6	862	21	US-09-791-537-116651

Sequence 26785, A
Sequence 121036, A
Sequence 44327, A
Sequence 118817, A
Sequence 125732, A
Sequence 116651, A

7	277	98.6	865	21	US-09-791-537-118818	Sequence 118818, A
8	277	98.6	866	11	US-08-750-458-2	Sequence 2, Appl1
9	277	98.6	866	21	US-09-791-537-44920	Sequence 44920, A
10	277	98.6	866	21	US-09-791-537-81933	Sequence 81933, A
11	277	98.6	876	18	US-09-487-558-44	Sequence 44, Appl
12	277	98.6	876	18	US-09-487-558B-44	Sequence 44, Appl
13	277	98.6	876	21	US-09-791-537-17502	Sequence 17502, A
14	277	98.6	876	21	US-09-791-537-47207	Sequence 47207, A
15	277	98.6	882	22	US-09-801-368-44	Sequence 44, Appl
16	277	98.6	882	22	US-09-801-368-44	Sequence 44, Appl
17	277	98.6	944	21	US-09-791-537-81931	Sequence 81931, A
18	277	98.6	971	21	US-09-791-537-137385	Sequence 137385, A
19	277	98.6	971	21	US-09-791-537-151258	Sequence 151258, A
20	269	95.7	956	21	US-09-791-537-118828	Sequence 118828, A
21	256	91.1	688	24	US-10-032-585-17876	Sequence 7876, Ap
22	256	91.1	719	16	US-09-248-796-17559	Sequence 17559, A
23	256	91.1	719	25	US-10-179-131-7398	Sequence 7398, Ap
24	256	91.1	719	27	US-09-409-409-17559	Sequence 17559, A
25	252	89.7	290	21	US-09-791-537-118820	Sequence 118820, A
26	252	89.7	727	21	US-09-791-537-148143	Sequence 148143, A
27	231	82.2	156	21	US-09-791-537-89028	Sequence 89028, A
28	231	82.2	510	21	US-09-791-537-118210	Sequence 118210, A
29	212	75.4	205	16	US-09-248-796-17564	Sequence 17564, A
30	212	75.4	205	27	US-09-409-409-17564	Sequence 17564, A
31	212	75.4	720	25	US-10-179-131-9251	Sequence 9251, Ap
32	210	74.7	551	19	US-09-538-092-448	Sequence 448, App
33	206	73.3	694	25	US-10-179-131-5254	Sequence 5254, Ap
34	204	72.6	135	21	US-09-791-537-79047	Sequence 79047, A
35	204	72.6	192	18	US-09-417-507-28191	Sequence 28191, A
36	204	72.6	269	21	US-09-791-537-7384	Sequence 7384, Ap
37	204	72.6	269	21	US-09-791-537-80768	Sequence 80768, A
38	204	72.6	298	18	US-09-487-558-230	Sequence 230, App
39	204	72.6	298	18	US-09-487-558B-230	Sequence 230, App
40	204	72.6	298	22	US-09-801-368-230	Sequence 230, App
41	201	71.5	730	18	US-09-487-558-126	Sequence 126, App
42	201	71.5	730	18	US-09-487-558B-126	Sequence 126, App
43	201	71.5	730	22	US-09-801-368-126	Sequence 126, App
44	195	69.4	383	21	US-09-791-537-46380	Sequence 46380, A
45	195	69.4	388	21	US-09-791-537-29921	Sequence 29921, A

ALIGNMENTS

RESULT 1
US-09-791-537-26785

; Sequence 26785, Application US/09791537
; GENERAL INFORMATION:
; APPLICANT: Bionomix, Inc.
; APPLICANT: Debe, Derek
; APPLICANT: Danzer, Joseph
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY-ME
; TITLE OF INVENTION: METHODS OF USE THEREOF
; FILE REFERENCE: 261/210
; CURRENT FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 153055
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 26785
; LENGTH: 1036
; TYPE: PRT
; ORGANISM: Neurospora crassa
US-09-791-537-26785

Query Match 100.0%; Score 281; DB 21; Length 1036;
Best Local Similarity 100.0%; Pred. No. 3.8e-27;
Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTNCFQTPTPLWRRNPDPGOLCNACGLFLKHLGVVRRPLSLKTDVTKKRR 50

Db 743 CTNCFQTPTPLWRRNPDPGOLCNACGLFLKHLGVVRRPLSLKTDVTKKRR 792

RESULT 2

US-09-791-537-121036
; Sequence 121036, Application US/09791537
; GENERAL INFORMATION:
; APPLICANT: Bionomix, Inc.
; APPLICANT: Debe, Derek
; APPLICANT: Danzer, Joseph
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY MEMBERS
; FILE REFERENCE: 261/210
; CURRENT APPLICATION NUMBER: US/09/791,537
; CURRENT FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 153055
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 121036
; LENGTH: 1036
; TYPE: PRT
; ORGANISM: Neurospora crassa
US-09-791-537-121036

Query Match 100.0%; Score 281; DB 21; Length 1036;
Best Local Similarity 100.0%; Pred. No. 3.8e-27;
Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTNCFQTTPPLWRNPDGQPLCNACGLFLKLGHVVRPLSLKTDVIKKNR 50
|||||
DB 743 CTNCFQTTPPLWRNPDGQPLCNACGLFLKLGHVVRPLSLKTDVIKKNR 792
|||||

RESULT 3

US-09-791-537-44327
; Sequence 44327, Application US/09791537
; GENERAL INFORMATION:
; APPLICANT: Bionomix, Inc.
; APPLICANT: Debe, Derek
; APPLICANT: Danzer, Joseph
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY MEMBERS
; FILE REFERENCE: 261/210
; CURRENT APPLICATION NUMBER: US/09/791,537
; CURRENT FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 153055
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 44327
; LENGTH: 66
; TYPE: PRT
; ORGANISM: pdb 4GATA
US-09-791-537-44327

Query Match 98.6%; Score 277; DB 21; Length 66;
Best Local Similarity 98.0%; Pred. No. 6.3e-28;
Matches 49; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTNCFQTTPPLWRNPDGQPLCNACGLFLKLGHVVRPLSLKTDVIKKNR 50
|||||
DB 12 CTNCFQTTPPLWRNPDGQPLCNACGLFLKLGHVVRPLSLKTDVIKKNR 61
|||||

RESULT 4

US-09-791-537-118817
; Sequence 118817, Application US/09791537
; GENERAL INFORMATION:
; APPLICANT: Bionomix, Inc.
; APPLICANT: Debe, Derek
; APPLICANT: Danzer, Joseph
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY MEMBERS
; FILE REFERENCE: 261/210
; CURRENT APPLICATION NUMBER: US/09/791,537
; CURRENT FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 153055
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 118817

; LENGTH: 725

; TYPE: PRT
; ORGANISM: Penicillium chrysogenum
US-09-791-537-118817

Query Match 98.6%; Score 277; DB 21; Length 725;
Best Local Similarity 98.0%; Pred. No. 8.6e-27;
Matches 49; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTNCFQTTPPLWRNPDGQPLCNACGLFLKLGHVVRPLSLKTDVIKKNR 50
|||||
DB 525 CTNCFQTTPPLWRNPDGQPLCNACGLFLKLGHVVRPLSLKTDVIKKNR 574
|||||

RESULT 5

US-09-791-537-125732
; Sequence 125732, Application US/09791537
; GENERAL INFORMATION:
; APPLICANT: Bionomix, Inc.
; APPLICANT: Debe, Derek
; APPLICANT: Danzer, Joseph
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY MEMBERS
; FILE REFERENCE: 261/210
; CURRENT APPLICATION NUMBER: US/09/791,537
; CURRENT FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 153055
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 125732
; LENGTH: 860
; TYPE: PRT
; ORGANISM: Penicillium roqueforti
US-09-791-537-125732

Query Match 98.6%; Score 277; DB 21; Length 860;
Best Local Similarity 98.0%; Pred. No. 1e-26;
Matches 49; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTNCFQTTPPLWRNPDGQPLCNACGLFLKLGHVVRPLSLKTDVIKKNR 50
|||||
DB 660 CTNCFQTTPPLWRNPDGQPLCNACGLFLKLGHVVRPLSLKTDVIKKNR 709
|||||

RESULT 6

US-09-791-537-116651
; Sequence 116651, Application US/09791537
; GENERAL INFORMATION:
; APPLICANT: Bionomix, Inc.
; APPLICANT: Debe, Derek
; APPLICANT: Danzer, Joseph
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY MEMBERS
; FILE REFERENCE: 261/210
; CURRENT APPLICATION NUMBER: US/09/791,537
; CURRENT FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 153055
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 116651
; LENGTH: 862
; TYPE: PRT
; ORGANISM: Penicillium chrysogenum
US-09-791-537-116651

Query Match 98.6%; Score 277; DB 21; Length 862;
Best Local Similarity 98.0%; Pred. No. 1e-26;
Matches 49; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTNCFQTTPPLWRNPDGQPLCNACGLFLKLGHVVRPLSLKTDVIKKNR 50
|||||
DB 662 CTNCFQTTPPLWRNPDGQPLCNACGLFLKLGHVVRPLSLKTDVIKKNR 711
|||||

RESULT 7

US-09-791-537-118818
; Sequence 118818, Application US/09791537
; GENERAL INFORMATION:
; APPLICANT: Bionomix, Inc.
; APPLICANT: Debe, Derek
; APPLICANT: Danzer, Joseph
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY MEMBERS
; FILE REFERENCE: 261/210
; CURRENT APPLICATION NUMBER: US/09/791,537
; CURRENT FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 153055
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 118818
; LENGTH: 865
; TYPE: PRT
; ORGANISM: Penicillium urticae
US-09-791-537-118818

Query Match 98.6%; Score 277; DB 21; Length 865;
Best Local Similarity 98.0%; Pred. No. 1e-26; 0; Indels 0; Gaps 0;
Matches 49; Conservative 1; Mismatches 0;

QY 1 CTNCFQTPTPLWRRNPQGQPLCNACGLFLKLHGVRPLSLKTDVIKRRNR 50
|||||
Db 665 CTNCFQTPTPLWRRNPQGQPLCNACGLFLKLHGVRPLSLKTDVIKRRNR 714
|||||

RESULT 8
US-08-750-458-2
; Sequence 2, Application US/08750458
; GENERAL INFORMATION:
; APPLICANT: Christensen, Tove
; APPLICANT: Hynes, Michael J.
; TITLE OF INVENTION: A Fungus Wherein The Area Gene Has Been Modified And An Area
; NUMBER OF SEQUENCE ADDRESSES: 2
; CORRESPONDENCE ADDRESSES:
; ADDRESS: Novo Nordisk of North America, Inc.
; STREET: 405 Lexington Avenue, Suite 6400
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10174-6401
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; FILING DATE: 3-December-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Rozek, Carol E.
; REGISTRATION NUMBER: 36,993
; REFERENCE/DOCKET NUMBER: 4129.204-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 867 0123
; TELEFAX: 212 867 0298
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 866 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-750-458-2

Query Match 98.6%; Score 277; DB 11; Length 866;
Best Local Similarity 98.0%; Pred. No. 1e-26; 0; Indels 0; Gaps 0;
Matches 49; Conservative 1; Mismatches 0;

QY 1 CTNCFQTPTPLWRRNPQGQPLCNACGLFLKLHGVRPLSLKTDVIKRRNR 50
|||||

Db 664 CTNCFQTPTPLWRRNPQGQPLCNACGLFLKLHGVRPLSLKTDVIKRRNR 713
|||||

RESULT 9
US-09-791-537-44920
; Sequence 44920, Application US/09791537
; GENERAL INFORMATION:
; APPLICANT: Bionomix, Inc.
; APPLICANT: Debe, Derek
; APPLICANT: Danzer, Joseph
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY ME
; FILE REFERENCE: 261/210
; CURRENT APPLICATION NUMBER: US/09/791,537
; CURRENT FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 153055
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 44920
; LENGTH: 866
; TYPE: PRT
; ORGANISM: Aspergillus parasiticus
US-09-791-537-44920

Query Match 98.6%; Score 277; DB 21; Length 866;
Best Local Similarity 98.0%; Pred. No. 1e-26; 0; Indels 0; Gaps 0;
Matches 49; Conservative 1; Mismatches 0;

QY 1 CTNCFQTPTPLWRRNPQGQPLCNACGLFLKLHGVRPLSLKTDVIKRRNR 50
|||||
Db 664 CTNCFQTPTPLWRRNPQGQPLCNACGLFLKLHGVRPLSLKTDVIKRRNR 713
|||||

RESULT 10
US-09-791-537-81933
; Sequence 81933, Application US/09791537
; GENERAL INFORMATION:
; APPLICANT: Bionomix, Inc.
; APPLICANT: Debe, Derek
; APPLICANT: Danzer, Joseph
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY ME
; FILE REFERENCE: 261/210
; CURRENT APPLICATION NUMBER: US/09/791,537
; CURRENT FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 153055
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 81933
; LENGTH: 866
; TYPE: PRT
; ORGANISM: Aspergillus oryzae
US-09-791-537-81933

Query Match 98.6%; Score 277; DB 21; Length 866;
Best Local Similarity 98.0%; Pred. No. 1e-26; 0; Indels 0; Gaps 0;
Matches 49; Conservative 1; Mismatches 0;

QY 1 CTNCFQTPTPLWRRNPQGQPLCNACGLFLKLHGVRPLSLKTDVIKRRNR 50
|||||
Db 664 CTNCFQTPTPLWRRNPQGQPLCNACGLFLKLHGVRPLSLKTDVIKRRNR 713
|||||

RESULT 11
US-09-487-558-44
; Sequence 44, Application US/09487558
; GENERAL INFORMATION:
; APPLICANT: Busby, Robert
; APPLICANT: Cali, Brian
; APPLICANT: Hecht, Peter
; APPLICANT: Holtzman, Doug
; APPLICANT: Madden, Kevin
; APPLICANT: Maxon, Mary
; APPLICANT: Milne, Todd
; APPLICANT: Norman, Thea

APPLICANT: Royer, John
APPLICANT: Salama, Sofie
APPLICANT: Sherman, Amir
APPLICANT: Silva, Jeff
APPLICANT: Summers, Eric
TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production
FILE REFERENCE: 109272.147
CURRENT APPLICATION NUMBER: US/09/487,558
PRIOR APPLICATION NUMBER: US/09/801,368
PRIOR FILING DATE: 2001-03-07
PRIOR APPLICATION NUMBER: US/09/487,558
PRIOR FILING DATE: 2000-01-19
PRIOR APPLICATION NUMBER: US/09/160,587
PRIOR FILING DATE: 1999-10-20
NUMBER OF SEQ ID NOS: 440
SOFTWARE: PatentIn version 3.0
SEQ ID NO 44
LENGTH: 876
TYPE: PRT
ORGANISM: Aspergillus nidulans
US-09-487-558-44

Query Match 98.6%; Score 277; DB 18; Length 876;
Best Local Similarity 98.0%; Pred. No. 1.1e-26;
Matches 49; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTNCFTQTTPLWRNPDGQPLCNACGLFLKLGWVRPLSLKTDVIKKNR 50
|||||
Db 673 CTNCFTQTTPLWRNPDGQPLCNACGLFLKLGWVRPLSLKTDVIKKNR 722

RESULT 12

US-09-487-558B-44
Sequence 44, Application US/09487558B

GENERAL INFORMATION:
APPLICANT: Busby, Robert
APPLICANT: Cali, Brian
APPLICANT: Hecht, Peter
APPLICANT: Holtzman, Doug
APPLICANT: Madden, Kevin
APPLICANT: Maxon, Mary
APPLICANT: Milne, Todd
APPLICANT: Norman, Thea
APPLICANT: Royer, John
APPLICANT: Salama, Sofie
APPLICANT: Sherman, Amir
APPLICANT: Silva, Jeff
APPLICANT: Summers, Eric
TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi
FILE REFERENCE: 109272.130
CURRENT APPLICATION NUMBER: US/09/487,558B
CURRENT FILING DATE: 2000-01-19
PRIOR APPLICATION NUMBER: US/09/487,558
PRIOR FILING DATE: 1999-10-20
NUMBER OF SEQ ID NOS: 446
SOFTWARE: PatentIn version 3.0
SEQ ID NO 44
LENGTH: 876
TYPE: PRT
ORGANISM: Aspergillus nidulans
US-09-487-558B-44

Query Match 98.6%; Score 277; DB 18; Length 876;
Best Local Similarity 98.0%; Pred. No. 1.1e-26;
Matches 49; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTNCFTQTTPLWRNPDGQPLCNACGLFLKLGWVRPLSLKTDVIKKNR 50
|||||
Db 673 CTNCFTQTTPLWRNPDGQPLCNACGLFLKLGWVRPLSLKTDVIKKNR 722

RESULT 13
US-09-791-537-17502
Sequence 17502, Application US/09791537
GENERAL INFORMATION:
APPLICANT: Bionomix, Inc.
APPLICANT: Debe, Derek
APPLICANT: Danzer, Joseph
TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY ME
TITLE OF INVENTION: METHODS OF USE THEREOF
FILE REFERENCE: 261/210
CURRENT APPLICATION NUMBER: US/09/791,537
CURRENT FILING DATE: 2001-02-22
NUMBER OF SEQ ID NOS: 153055
SOFTWARE: PatentIn version 3.0
SEQ ID NO 17502
LENGTH: 876
TYPE: PRT
ORGANISM: Emmericella nidulans
US-09-791-537-17502

Query Match 98.6%; Score 277; DB 21; Length 876;
Best Local Similarity 98.0%; Pred. No. 1.1e-26;
Matches 49; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTNCFTQTTPLWRNPDGQPLCNACGLFLKLGWVRPLSLKTDVIKKNR 50
|||||
Db 673 CTNCFTQTTPLWRNPDGQPLCNACGLFLKLGWVRPLSLKTDVIKKNR 722

RESULT 14

US-09-791-537-47207
Sequence 47207, Application US/09791537

GENERAL INFORMATION:
APPLICANT: Bionomix, Inc.
APPLICANT: Debe, Derek
APPLICANT: Danzer, Joseph
TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY ME
TITLE OF INVENTION: METHODS OF USE THEREOF
FILE REFERENCE: 261/210
CURRENT APPLICATION NUMBER: US/09/791,537
CURRENT FILING DATE: 2001-02-22
NUMBER OF SEQ ID NOS: 153055
SOFTWARE: PatentIn version 3.0
SEQ ID NO 47207
LENGTH: 876
TYPE: PRT
ORGANISM: Emmericella nidulans
US-09-791-537-47207

Query Match 98.6%; Score 277; DB 21; Length 876;
Best Local Similarity 98.0%; Pred. No. 1.1e-26;
Matches 49; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTNCFTQTTPLWRNPDGQPLCNACGLFLKLGWVRPLSLKTDVIKKNR 50
|||||
Db 673 CTNCFTQTTPLWRNPDGQPLCNACGLFLKLGWVRPLSLKTDVIKKNR 722

RESULT 15

US-09-801-368-44
Sequence 44, Application US/09801368

GENERAL INFORMATION:
APPLICANT: Busby, Robert
APPLICANT: Cali, Brian
APPLICANT: Hecht, Peter
APPLICANT: Holtzman, Doug
APPLICANT: Madden, Kevin
APPLICANT: Maxon, Mary
APPLICANT: Milne, Todd
APPLICANT: Norman, Thea
APPLICANT: Royer, John
APPLICANT: Salama, Sofie
APPLICANT: Sherman, Amir

; APPLICANT: Silva, Jeff
; APPLICANT: Summers, Eric
; TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi
; FILE REFERENCE: 109272.147
; CURRENT APPLICATION NUMBER: US/09/801,368
; CURRENT FILING DATE: 2001-03-07
; PRIOR APPLICATION NUMBER: US 09/487,558
; PRIOR FILING DATE: 2000-01-19
; PRIOR APPLICATION NUMBER: US 60/160,587
; PRIOR FILING DATE: 1999-10-20
; NUMBER OF SEQ ID NOS: 440
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 44
; LENGTH: 876
; TYPE: PRT
; ORGANISM: Aspergillus nidulans
US-09-801-368-44

Query Match 98.6%; Score 277; DB 22; Length 876;
Best Local Similarity 98.0%; Pred. No. 1.1e-26;
Matches 49; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CTNCFQTTPPLWRNPDGQPLCNACGLFLKHLGVWRPLSLKTDVIKKRNR 50
|||||
Db 673 CTNCFQTTPPLWRNPDGQPLCNACGLFLKHLGVWRPLSLKTDVIKKRNR 722
|||||

Search completed: January 3, 2003, 19:59:15
Job time : 188.5 secs



GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 3, 2003, 19:50:43 ; Search time 19.5 seconds
(without alignments)
181.016 Million cell updates/sec

Title: US-09-725-010-3
Perfect score: 281
Sequence: 1 CTNCFQTTPWRRNPDPGQP.....LHGVRPLSLKTDVKKRNR 50

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 258925 seqs, 70596210 residues

Total number of hits satisfying chosen parameters: 258925

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending_Patents_AA_New.*
1: /cgn2_6/ptodata/2/paa/PCT_NEW_COMB.pep.*
2: /cgn2_6/ptodata/2/paa/US06_NEW_COMB.pep.*
3: /cgn2_6/ptodata/2/paa/US07_NEW_COMB.pep.*
4: /cgn2_6/ptodata/2/paa/US08_NEW_COMB.pep.*
5: /cgn2_6/ptodata/2/paa/US09_NEW_COMB.pep.*
6: /cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep.*
7: /cgn2_6/ptodata/2/paa/US60_NEW_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	281	100.0	50	5	US-09-725-010-3
2	256	91.1	50	5	US-09-725-010-2
3	188	66.9	347	5	US-09-724-676-87510
4	188	66.9	347	5	US-09-724-676-87510
5	182	64.8	357	5	US-09-724-676-92485
6	182	64.8	357	5	US-09-724-676-92485
7	178	63.3	455	5	US-09-724-676-92294
8	178	63.3	455	5	US-09-724-676-92294
9	129	45.9	203	5	US-09-724-676-87508
10	129	45.9	203	5	US-09-724-676-87508
11	129	45.9	471	5	US-09-724-676-87509
12	129	45.9	471	5	US-09-724-676-87509
13	100	35.6	291	6	US-10-290-627-4
14	100	35.6	291	6	US-10-286-264-98
15	89.5	31.9	139	6	US-10-278-536-32
16	89.5	31.9	139	6	US-10-171-404A-28
17	66	23.5	593	1	PCT-US02-38445-17
18	62.5	22.2	267	5	US-09-724-676-61656
19	62.5	22.2	267	5	US-09-724-676-61656
20	62.5	22.2	270	5	US-09-724-676-61641
21	62.5	22.2	270	5	US-09-724-676-61641
22	62.5	22.2	301	5	US-09-724-676-61710
23	62.5	22.2	301	5	US-09-724-676-61710
24	62.5	22.2	304	5	US-09-724-676-61679
25	62.5	22.2	304	5	US-09-724-676-61679
26	62.5	22.2	372	5	US-09-724-676-61663

ALIGNMENTS

RESULT 1

US-09-725-010-3
; Sequence 3, Application US/09725010
; GENERAL INFORMATION:

; APPLICANT: SUNSTROM, PAULA
; TITLE OF INVENTION: METHODS FOR ALTERING THE EXPRESSION OF HYPHAL-SPECIFIC
; FILE REFERENCE: 23878.0005
; CURRENT APPLICATION NUMBER: US/09725.010
; CURRENT FILING DATE: 2000-11-29
; PRIOR APPLICATION NUMBER: 60/167,672
; PRIOR FILING DATE: 1999-11-29
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 50
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-725-010-3

Query Match 100.0%; Score 281; DB 5; Length 50;
Best Local Similarity 100.0%; Pred. No. 1.3e-27; Mismatches 0; Indels 0; Gaps 0;
Matches 50; Conservative 0;

Qy 1 CTNCFQTTPWRRNPDPGQPLCNAGLFLKLGVRPLSLKTDVKKRNR 50
|||||
Db 1 CTNCFQTTPWRRNPDPGQPLCNAGLFLKLGVRPLSLKTDVKKRNR 50

RESULT 2

US-09-725-010-2
; Sequence 2, Application US/09725010
; GENERAL INFORMATION:

; APPLICANT: SUNSTROM, PAULA
; TITLE OF INVENTION: METHODS FOR ALTERING THE EXPRESSION OF HYPHAL-SPECIFIC
; FILE REFERENCE: 23878.0005
; CURRENT APPLICATION NUMBER: US/09725.010
; CURRENT FILING DATE: 2000-11-29
; PRIOR APPLICATION NUMBER: 60/167,672
; PRIOR FILING DATE: 1999-11-29
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 50
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-725-010-2

RESULT 11
US-09-724-676-87509

; APPLICANT: Riechmann, Jose-Luis

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 3, 2003, 19:51:53 ; Search time 21.5 seconds
(without alignments)
44.073 Million cell updates/sec

Title: US-09-725-010-3
Perfect score: 281
Sequence: 1 CTNCFQTTPRLRRNPDPG.....LHGVRPLSLKTDVIKRRNR 50

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 117078 seqs, 18951520 residues

Total number of hits satisfying chosen parameters: 117078

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published_Applications_AA.*
1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	277	98.6	876	10	US-09-801-368-44
2	204	72.6	298	10	US-09-801-368-230
3	201	71.5	730	10	US-09-801-368-126
4	183	65.1	442	10	US-09-749-728B-11
5	179	63.7	532	10	US-09-801-368-360
6	178	63.3	549	10	US-09-801-368-358
7	158	56.2	128	10	US-09-925-300-1573
8	93	33.1	359	9	US-10-029-180-68
9	93	33.1	440	9	US-10-029-180-124
10	93	33.1	451	9	US-10-029-180-123
11	75	26.7	588	10	US-09-801-368-46
12	56.5	20.1	136	10	US-09-948-018-6
13	56.5	20.1	355	10	US-09-948-018-2
14	56.5	20.1	380	10	US-09-948-018-36
15	56.5	20.1	1194	9	US-09-738-626-5832
16	54	19.2	74	10	US-09-764-846-181
17	53	18.9	108	10	US-09-799-946-11
18	53	18.9	387	10	US-09-967-552A-36
19	52.5	18.7	2237	12	US-10-033-026-8

20	51	18.1	292	10	US-09-815-242-13782
21	51	18.1	1146	10	US-09-824-734-2
22	51	18.1	1148	10	US-09-900-237-4
23	50.5	18.0	315	10	US-09-808-387-26
24	50.5	18.0	373	10	US-09-808-387-24
25	50.5	18.0	380	10	US-09-729-674-128
26	50.5	18.0	386	10	US-09-808-387-6
27	50.5	18.0	555	10	US-09-808-387-22
28	50.5	18.0	557	10	US-09-808-387-4
29	50.5	18.0	574	10	US-09-808-387-2
30	49.5	17.6	137	10	US-09-508-711-101
31	49.5	17.6	188	10	US-09-764-864-1424
32	49.5	17.6	188	10	US-09-764-877-1233
33	49.5	17.6	338	10	US-09-764-864-948
34	49.5	17.6	3712	9	US-10-108-605-103
35	49	17.4	87	10	US-09-867-550-1118
36	49	17.4	252	10	US-09-816-669A-12
37	48.5	17.3	314	10	US-09-880-705-2
38	48.5	17.3	415	12	US-10-005-947-2
39	48.5	17.3	2502	10	US-09-772-316-1
40	48.5	17.3	3084	10	US-09-938-275-4
41	48	17.1	415	10	US-09-925-300-1352
42	48	17.1	724	10	US-09-759-010-6
43	48	17.1	732	9	US-09-991-496-18
44	48	17.1	732	10	US-09-759-010-5
45	48	17.1	732	10	US-09-874-923-18

ALIGNMENTS

RESULT 1
US-09-801-368-44
; Sequence 44, Application US/09801368
; Patent No. US20020128250A1
; GENERAL INFORMATION:
; APPLICANT: Busby, Robert
; APPLICANT: Cali, Brian
; APPLICANT: Hecht, Peter
; APPLICANT: Holtzman, Doug
; APPLICANT: Madden, Kevin
; APPLICANT: Maxon, Mary
; APPLICANT: Milne, Todd
; APPLICANT: No. US20020128250Alman, Thea
; APPLICANT: Royer, John
; APPLICANT: Salama, Sofie
; APPLICANT: Sherman, Amir
; APPLICANT: Silva, Jeff
; APPLICANT: Summers, Eric
; TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi
; FILE REFERENCE: 109272.147
; CURRENT APPLICATION NUMBER: US/09/801,368
; PRIOR FILING DATE: 2001-03-07
; PRIOR APPLICATION NUMBER: US 09/487,558
; PRIOR FILING DATE: 2000-01-19
; PRIOR APPLICATION NUMBER: US 60/160,587
; NUMBER OF SEQ ID NOS: 440
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 44
; LENGTH: 876
; TYPE: PRT
; ORGANISM: Aspergillus nidulans
US-09-801-368-44

Query Match 98.6%; Score 277; DB 10; Length 876;
Best Local Similarity 98.0%; Pred. No. 7.9e-29;
Matches 49; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 CTNCFQTTPRLRRNPDPGCPICNACGLFLKLGVRPLSLKTDVIKRRNR 50
DB 673 CTNCFQTTPRLRRNPDPGCPICNACGLFLKLGVRPLSLKTDVIKRRNR 722

; CURRENT APPLICATION NUMBER: US/09/801.368
; CURRENT FILING DATE: 2001-03-07
; PRIOR APPLICATION NUMBER: US 09/487,558
; PRIOR FILING DATE: 2000-01-19
; PRIOR APPLICATION NUMBER: US 60/160,587
; PRIOR FILING DATE: 1999-10-20
; NUMBER OF SEQ ID NOS: 440
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 360
; LENGTH: 532
; TYPE: PRT
; ORGANISM: Penicillium chrysogenum
US-09-801-368-360

Query Match 63.7%; Score 179; DB 10; Length 532;
Best Local Similarity 58.0%; Pred. No. 4.4e-16;
Matches 29; Conservative 7; Mismatches 14; Indels 0; Gaps 0;

Qy 1 CTNCFQTPTPLWRRNPDGQPLCNACGLFLKLGHVVRPLSLKTDVKKRNR 50
||| ||||| :|:|||||: |||| || :||:|
Db 238 CQCGTTVTPLWRRDEQHPHCNACGLYKLGCGYRTNMKKSIIKKRR 287

RESULT 6
US-09-801-368-358
; Sequence 358, Application US/09801368
; Patent No. US20020128250A1
; GENERAL INFORMATION:
; APPLICANT: Busdy, Robert
; APPLICANT: Cali, Brian
; APPLICANT: Hecht, Peter
; APPLICANT: Holtzman, Doug
; APPLICANT: Madden, Kevin
; APPLICANT: Maxon, Mary
; APPLICANT: Milne, Todd
; APPLICANT: No. US20020128250A1man, Thea
; APPLICANT: Royer, John
; APPLICANT: Salama, Sofie
; APPLICANT: Sherman, Amir
; APPLICANT: Silva, Jeff
; APPLICANT: Summers, Eric
; TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi
; FILE REFERENCE: 109272.147
; CURRENT APPLICATION NUMBER: US/09/801.368
; CURRENT FILING DATE: 2001-03-07
; PRIOR APPLICATION NUMBER: US 09/487,558
; PRIOR FILING DATE: 2000-01-19
; PRIOR APPLICATION NUMBER: US 60/160,587
; PRIOR FILING DATE: 1999-10-20
; NUMBER OF SEQ ID NOS: 440
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 358
; LENGTH: 549
; TYPE: PRT
; ORGANISM: Aspergillus nidulans
US-09-801-368-358

Query Match 63.3%; Score 178; DB 10; Length 549;
Best Local Similarity 58.0%; Pred. No. 6.2e-16;
Matches 29; Conservative 8; Mismatches 13; Indels 0; Gaps 0;

Qy 1 CTNCFQTPTPLWRRNPDGQPLCNACGLFLKLGHVVRPLSLKTDVKKRNR 50
||| ||||| :|:|||||: |||| || :||:|
Db 251 CQCGTTVTPLWRRDEQHPHCNACGLYKLGHSYRPTTMKKTIIKKRR 300

RESULT 7
US-09-925-300-1573
; Sequence 1573, Application US/09925300
; Patent No. US20020151681A1
; GENERAL INFORMATION:
; APPLICANT: Craig Rosen,
; APPLICANT: Steve Ruben

; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA101
; CURRENT APPLICATION NUMBER: US/09/925,300
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05988
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1890
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1573
; LENGTH: 128
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-925-300-1573

Query Match 56.2%; Score 158; DB 10; Length 128;
Best Local Similarity 62.2%; Pred. No. 5.3e-14;
Matches 28; Conservative 7; Mismatches 10; Indels 0; Gaps 0;

Qy 6 TOTTPLWRRNPDGQPLCNACGLFLKLGHVVRPLSLKTDVKKRNR 50
||| ||||| :|:|||||: |||| || :||:|
Db 2 TTTTLWRRNANGDPVCNACGLYKLVNVRPLTMKKGIOQRNR 46

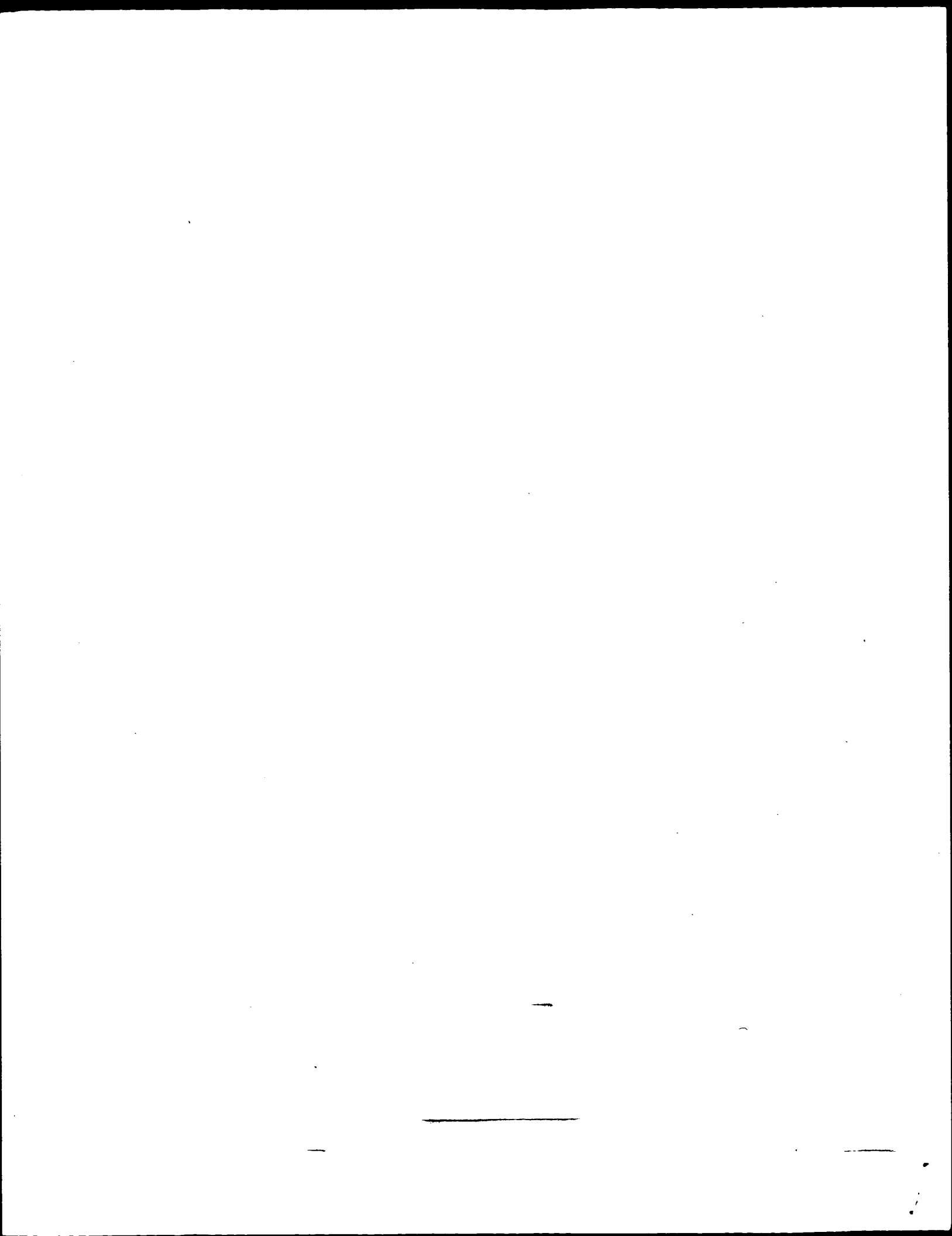
RESULT 8
US-10-029-180-68
; Sequence 68, Application US/10029180
; Publication No. US20020182708A1
; GENERAL INFORMATION:
; APPLICANT: Cali, Brian M.
; APPLICANT: Holtzman, Doug
; APPLICANT: Madden, Kevin T.
; APPLICANT: Milna, G. Todd
; APPLICANT: Sherman, Amir
; APPLICANT: Silva, Jeffrey C.
; APPLICANT: Trueheart, Josh
; APPLICANT: Zhang, Lixin
; TITLE OF INVENTION: No. US20020182708A1el Regulators of Fungal Gene Expression
; FILE REFERENCE: MIC-004
; CURRENT APPLICATION NUMBER: US/10/029,180
; CURRENT FILING DATE: 2001-12-22
; PRIOR APPLICATION NUMBER: US 60/257,431
; PRIOR FILING DATE: 2000-12-22
; NUMBER OF SEQ ID NOS: 138
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 68
; LENGTH: 359
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: fungal gene
US-10-029-180-68

Query Match 33.1%; Score 93; DB 9; Length 359;
Best Local Similarity 57.6%; Pred. No. 7.2e-05;
Matches 19; Conservative 3; Mismatches 9; Indels 2; Gaps 2;

Qy 1 CTNCFQTPTPLWRRNPDG-QPLCNACGL-FLKL 31
|:| |||||: |||||: ||
Db 308 CHCNSRSETPWRRGPDGPRTLNACGLHYAKL 340

RESULT 9
US-10-029-180-124
; Sequence 124, Application US/10029180
; Publication No. US20020182708A1
; GENERAL INFORMATION:
; APPLICANT: Cali, Brian M.
; APPLICANT: Holtzman, Doug
; APPLICANT: Madden, Kevin T.
; APPLICANT: Milna, G. Todd
; APPLICANT: Sherman, Amir

US-09-948-018-2
; Sequence 2, Application US/09948018
; Patent No. US20020150977A1



GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 3, 2003, 18:48:02 ; Search time 44.5 Seconds
(without alignments)
149.720 Million cell updates/sec

Title: US-09-725-010-3

Perfect score: 281

Sequence: 1 CTNCFQTTPWRRPDGQP.....LHGVRPLSLKTDVTKRNR 50

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_101002:*

- 1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
- 2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
- 3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
- 4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
- 5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
- 6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
- 7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
- 8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
- 9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
- 10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
- 11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
- 12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
- 13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
- 14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
- 15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
- 16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
- 17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
- 18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
- 19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
- 20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
- 21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
- 22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
- 23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	281	100.0	50	22	AA82413
2	277	98.6	853	20	AAW81079
3	277	98.6	866	17	AA88408
4	277	98.6	866	18	AAW31630
5	256	91.1	50	22	AA82414
6	190	67.6	950	22	ABB71271
7	188	66.9	279	22	ABG74577
8	188	66.9	279	23	ABP41885
9	188	66.9	449	19	AAW56703
10	183	65.1	442	22	AA899920

11	183	65.1	442	22	AA864319	Human GATA4 protei
12	183	65.1	442	22	AA864849	Heart muscle cell
13	180	64.1	486	22	ABB64550	Drosophila melanog
14	178	63.3	443	23	AAU79800	Mouse GATA-3 trans
15	178	63.3	443	23	AAU79801	Transgenic animal
16	178	63.3	443	23	ABB85021	Pain regulated pro
17	178	63.3	443	23	ABB85022	Pain regulated pro
18	178	63.3	444	20	AA423279	GATA-3 amino acid
19	176	62.6	734	22	ABB71878	Drosophila melanog
20	158	56.2	128	21	AA856995	Human prostate can
21	139.5	49.6	785	22	AA892512	Human protein sequ
22	139.5	49.6	1281	21	AA800190	Breast cancer prot
23	139.5	49.6	1281	22	AAE02189	Human breast cance
24	128.5	45.7	1294	21	AA840550	Human ORFX ORF314
25	117	41.6	273	22	AA878812	Zinc finger protei
26	108.5	38.6	187	21	AA815686	Arabidopsis thalia
27	108.5	38.6	271	21	AA815685	Arabidopsis thalia
28	108.5	38.6	309	21	AA815684	Arabidopsis thalia
29	108	38.4	119	22	AA825514	Human protein sequ
30	107	38.1	210	21	AA841113	Human ORFX ORF877
31	100.5	35.8	276	21	AA808697	Arabidopsis thalia
32	100.5	35.8	276	21	AA842639	Arabidopsis thalia
33	100.5	35.8	294	21	AA808696	Arabidopsis thalia
34	100.5	35.8	294	21	AA842638	Arabidopsis thalia
35	100.5	35.8	295	21	AA808695	Arabidopsis thalia
36	100.5	35.8	295	21	AA842637	Arabidopsis thalia
37	100	35.6	163	21	AA833111	Pinus radiata tran
38	97	34.5	264	23	AAU93001	Arabidopsis trans
39	96.5	34.3	278	21	AA811326	Arabidopsis thalia
40	96.5	34.3	285	21	AA821607	Arabidopsis thalia
41	96.5	34.3	287	21	AA811325	Arabidopsis thalia
42	96.5	34.3	295	21	AA821606	Arabidopsis thalia
43	96.5	34.3	297	21	AA811324	Arabidopsis thalia
44	96.5	34.3	329	21	AA821605	Arabidopsis thalia
45	96	34.2	240	23	AAU93121	Arabidopsis trans

ALIGNMENTS

RESULT 1
AA82413
ID AA82413 standard; Protein; 50 AA.
XX
AC AA82413;
XX
DT 06-AUG-2001 (first entry)
DE NIT2 DNA binding domain.
XX
DE
XX
KW Hyphal wall protein; HWP1; hyphal-specific gene; infection;
KW candidiasis; therapy; antifungal; fungicide; transcription factor;
KW NIT2; DNA binding protein.
XX
OS Neurospora crassa.
XX
PN WO200138550-A2.
XX
PD 31-MAY-2001.
XX
PF 29-NOV-2000; 2000WO-US32464.
XX
PR 29-NOV-1999; 99US-0167672.
XX
PA (SUND/) SUNDSTROM P.
XX
PI Sundstrom P;
XX
DR WPI; 2001-367698/38.
XX
PT Interfering with expression of hyphal-specific genes in fungus for
PT inhibiting fungal cell growth involves interfering with transcription
PT of hyphal-specific genes mediated by cis acting sequences

Claim 31; Page 80; 95pp; English.

The present sequence is that of the NIT2 DNA binding domain of *Neurospora crassa*. The NIT2 protein is a global positive-acting transcription factor of nitrogen structural genes when preferred N-sources are lacking. The promoter region (see AAR90471) of the *Candida albicans* hyphal-specific Hwpl gene contains numerous binding sites for NIT2. GAR99 (see AAB82414), showing 92% homology to NIT2, was identified in *C. albicans*, suggesting that regulation of Hwpl expression may be tied to nitrogen regulatory events. The Hwpl gene encodes an adhesion required for the pathogenesis of candidiasis. The invention provides a method for interfering with the expression of hyphal-specific genes in a fungus resulting in inhibition of cell growth. This involves interfering with the transcription of the hyphal-specific gene mediated by cis-acting sequences. The interfering step may involve manipulating the binding of a DNA binding protein to a cis-regulatory element. The fungus may be a non-pathogenic or a pathogenic fungus, such as *C. albicans*, and the hyphal-specific gene may be Hwpl, in which the NIT2 binding sites act as the cis-regulatory elements. The method provides a means of treating infection of mammalian hosts such as immunocompromised or immunosuppressed humans, including those having AIDS or undergoing transplantation or anti-cancer therapy, burns patients, patients with diabetic ketoacidosis, and patients in which the normal microbial flora has been disrupted because of disease, trauma or chemical, radiation or other immunosuppressive prophylaxis. Also provided is a method for characterising genes under control of a DNA binding protein.

PT - useful for hydrolyzing protein-containing materials, and producing
PT food products with improved organoleptic properties
XX
XX Claim 11; Pages 28-30; 53pp; English.
XX
CC This is the amino acid sequence of the area protein of *Aspergillus*
CC oryzae used in the method of the invention involving the doubled
CC expression of peptidases in a new *Koji* mold. The *Koji* mold is
CC useful for hydrolyzing protein-containing materials, and for
CC over-producing proteolytic enzymes. When hydrolyzing
CC protein-containing material (containing 5 mM L-glutamine), the
CC mold may be used in combination with an enzyme and/or a
CC microorganism providing a prolidase activity. Additionally, the
CC mold is useful in fermentation processes for creating food
CC products. The *Koji* mold is used to produce high levels of
CC endopeptidases and exopeptidases, including enhanced
CC prolyl-dipeptidyl-peptidase activity, which may be useful for
CC releasing L-glutamine from peptides.

RESULT 3	
AAR88408	
ID	AAR88408 standard; Protein; 866 AA.
XX	
AC	AAR88408;
XX	
DT	26-JUN-1996 (first entry)
XX	
DE	AreaA activator protein.
XX	
KW	AreaA activator; protease-free; host cell.
XX	
OS	<u>Aspergillus oryzae strain IFO4177.</u>
XX	
PN	WO9535385-A1.
XX	
PD	28-DEC-1995.
XX	
PF	19-JUN-1995; 95WO-DK00254.
XX	
PR	17-JUN-1994; 94DK-0000717.
XX	
PA	(NOVO) NOVO-NORDISK AS.
XX	
PI	Christensen T, Hynes MJ;
XX	
DR	WPI; 1996-058422/06.
XX	
DR	N-PSDB; AAT09946.
XX	
PT	Fungus with a modified areaA gene - useful to produce proteins
XX	
PT	susceptible to degradation by protease(s)
XX	
PS	Claim 23; Page 23-25; 42pp; English.
XX	
CC	The AreaA activator protein (AAR88408) of <i>Aspergillus oryzae</i> IFO4177
CC	controls nitrogen catabolism and influences the production of
CC	extracellular proteases. It is the product of the areaA gene
CC	(AAT09946). Inactivation of the activator, by deletion or
CC	antisense inhibition of the gene, yields protease-free
CC	<i>Aspergillus</i> cells, e.g. strain T0C913, suitable as hosts for
CC	prodn. of protease-susceptible heterologous proteins such as
XX	enzymes and industrial or therapeutic proteins.
XX	

QY	1	CTNCFQTTPLRNRNDGQPLCNACGLFLKLHGVRP ^L SLKTDVIKKRNR	50
		:	
Dp	664	CTNCFQTTPLRNRPEGQPLCNACGLFLKLHGVRP ^L SLKTDVIKKRNR	713

1 CTNCFQTTPLRNRPDGQPLCNACGLFLKLHGVRPLSLKTDVIKKNR 50

XX Differentiation; heart muscle cell; cytokine; transcription factor;
 KW proliferation; surface antigen; heart disease; cardiomyocyte;
 KW bone marrow; umbilical blood cell; heart muscle degeneration;
 KW myocardial infarction.
 XX
 OS Homo sapiens.
 XX
 PN WO200148150-A1.
 XX
 PD 05-JUL-2001.
 XX
 PF 02-NOV-2000; 2000WO-JP07741.
 XX
 PR 28-DEC-1999; 99JP-0372826.
 XX
 PR 28-FEB-2000; 2000WO-JP01148.
 XX
 PA (KYOW) KYOWA HAKKO KOGYO KK.
 XX
 XX Umezawa A, Hata J, Fukuda K, Ogawa S, Sakurada K, Gojo S;
 PI Yamada Y;
 XX WPI; 2001-425655/45.
 XX N-PSDB; AAH4356.
 DR
 DR
 XX
 XX Cells capable of differentiating into cardiomyocytes and originating in
 PT bone marrow or umbilical blood cells for study of cardiomyocyte
 PT differentiation and treatment of heart disease .
 XX
 PS Claim 41; Page 98-99; 187pp; Japanese.
 XX
 CC The present invention describes cells originating in bone marrow or
 CC umbilical blood cells which are capable of differentiating into
 CC cardiomyocytes. Also described are: (1) cardiomyocytes produced by the
 CC differentiation of the cells; (2) a method for carrying out the
 CC differentiation into cardiomyocytes, regulated by a promotional and/or
 CC inhibitory factor; (3) a method for the differentiation of the cells
 CC into cell types other than cardiomyocytes; (4) drug compositions
 CC promoting the formation of heart muscle and regeneration of heart tissue
 CC which contain the cells; (5) a method for the production of antibodies
 CC which recognise the cells, especially antibodies which recognise a
 CC surface antigen on the cells; (6) a method for screening factors which
 CC promote the proliferation of the cells; (7) a method for immortalising
 CC the cells by expressing telomerase in them; (8) drug compositions for
 CC the treatment of heart disease which contain the immortalised cells; and
 CC (9) cell-free supernatant from the culture of the cells and its use in
 CC promoting their differentiation into cardiomyocytes. The cells are used
 CC in the treatment of diseases involving heart muscle degeneration, such
 CC as myocardial infarction and in the study of cardiomyocyte
 CC differentiation. AAH44351 to AAH44409 and AAB99915 to AAB99935 represent
 CC sequences used in the exemplification of the present invention.
 XX
 SQ Sequence 442 AA;
 Query Match 65.1%; Score 183; DB 22; Length 442;
 Best Local Similarity 60.0%; Pred. No. 1.4e-15;
 Matches 30; Conservative 10; Mismatches 10; Indels 0; Gaps 0;
 QY 1 CTNCFQTTPLRNPDGQPLNACGLFLKLGWVRPLSLKTDVIKKNR 50
 Db 271 CANGQTTTTLWRNABGEPCVNCAGLYMKLHGVRPLAMRKEGIQTRK 320
 RESULT 11
 AAG64319
 ID AAG64319 standard; Protein; 442 AA.
 XX
 AC AAG64319;
 XX
 XX 24-SEP-2001 (first entry)
 DT Human GATA4 protein.
 DE
 DE
 XX

KW Angiogenesis; cardiant; cell differentiating agent; bone marrow;
 KW heart muscle cell; heart disease; human; GATA4.
 XX
 OS Homo sapiens.
 XX
 PN WO200148149-A1.
 XX
 PD 05-JUL-2001.
 XX
 PF 28-FEB-2000; 2000WO-JP01148.
 XX
 PR 28-DEC-1999; 99JP-0372826.
 XX
 PR (KYOW) KYOWA HAKKO KOGYO KK.
 XX
 XX Umezawa A, Hata J, Fukuda K, Ogawa S, Sakurada K;
 PI WPI; 2001-418252/44.
 XX N-PSDB; AAH49591.
 DR
 DR
 XX
 XX New adult bone marrow-originated cells capable of differentiating into
 PT heart muscle cells, applicable as remedies for various heart diseases
 PT particularly with damaged heart muscle accompanying degeneration .
 XX
 PS Claim 23; Pages 74-76; 158pp; Japanese.
 XX
 CC The present invention relates to cells isolated from bone marrow, which
 CC are capable of at least differentiating into heart muscle cells. The
 CC cells are applicable as remedies for various heart diseases particularly
 CC with damaged heart muscle accompanying degeneration. The present sequence
 CC was used to illustrate the present invention.
 XX
 SQ Sequence 442 AA;
 Query Match 65.1%; Score 183; DB 22; Length 442;
 Best Local Similarity 60.0%; Pred. No. 1.4e-15;
 Matches 30; Conservative 10; Mismatches 10; Indels 0; Gaps 0;
 QY 1 CTNCFQTTPLRNPDGQPLNACGLFLKLGWVRPLSLKTDVIKKNR 50
 Db 271 CANGQTTTTLWRNABGEPCVNCAGLYMKLHGVRPLAMRKEGIQTRK 320
 RESULT 12
 AAG64849
 ID AAG64849 standard; Protein; 442 AA.
 XX
 AC AAG64849;
 XX
 XX 21-SEP-2001 (first entry)
 DT Heart muscle cell differentiation related protein SEQ ID NO: 11.
 DE
 DE
 XX
 KW Heart muscle cell; human; cell differentiation; heart disease.
 XX
 OS Homo sapiens.
 XX
 PN WO200148151-A1.
 XX
 PD 05-JUL-2001.
 XX
 PF 27-DEC-2000; 2000WO-JP09323.
 XX
 PR 28-DEC-1999; 99JP-0372826.
 XX
 PR 28-FEB-2000; 2000WO-JP01148.
 XX
 PR 02-NOV-2000; 2000WO-JP07741.
 XX
 XX (KYOW) KYOWA HAKKO KOGYO KK.
 XX
 XX Umezawa A, Hata J, Fukuda K, Ogawa S, Sakurada K, Gojo S;
 PI Yamada Y;
 XX WPI; 2001-425656/45.
 DR

```

CC      at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ      Sequence      486 AA;

Query Match      64.1%; Score 180; DB 22; Length 486;
Best Local Similarity 62.0%; Pred. No. 3.7e-15;
Matches 31; Conservative 7; Mismatches 12; Indels 0; Gaps 0;

QY      1 CTNCFQTPTPLWRRNPDPGLCNACGLFLKLHGVVRPLSLKTDVIKKNR 50
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      321 CANKCTTTTLLWRRNAGPVCNACGLYKLLHNVRPLTMKKEGIQTRNR 370

RESULT 14
AAU97800
ID      AAU97800 standard; Protein; 443 AA.
XX
AC      AAU97800;
XX
DT      13-AUG-2002 (first entry)
XX
DE      Mouse GATA-3 transcription factor.
XX
KW      Mouse; GATA-3 transcription factor; qgata-3; autoimmune disease;
KW      allergic disease; transgenic; immunosuppressive; antiallergic.
XX
OS      Mus sp.
XX
PN      JP2002112781-A.
XX      16-APR-2002.
PD
XX
XX      04-OCT-2000; 2000JP-0305252.
XX
XX      04-OCT-2000; 2000JP-0305252.
XX
XX      (SUMU ) SUMITOMO SEIYAKU KK.
XX
XX      WPI; 2002-447980/48.
XX
XX      N-PSDB; ABK50690.
XX
XX      Transgenic animal selectively expressing GATA-3, a recombinant vector
XX      selectively expressing GATA-3 of a nonhuman mammalian cell, and a
XX      preventive and/or treating agent for immune diseases -
XX
XX      Disclosure; Page 5-7; 8pp; Japanese.
XX
XX      The present invention relates to a recombinant vector for selectively
XX      expressing GATA-3 in a nonhuman mammalian cell, and which contains a
XX      DNA encoding qgata-3 or its mutant gene and is recombinated together
XX      with a promoter/enhancer sequence so that it may be expressed at a
XX      specified site. The compound is useful as a preventive and/or treating
XX      agent for autoimmune diseases and allergic diseases. The recombinant
XX      vector may be used to generate a transgenic mouse highly expressing
XX      mouse GATA-3. The present sequence represents mouse GATA-3 transcription
XX      factor.
XX
XX      Sequence      443 AA;

Query Match      63.3%; Score 178; DB 23; Length 443;
Best Local Similarity 60.0%; Pred. No. 6.1e-15;
Matches 30; Conservative 8; Mismatches 12; Indels 0; Gaps 0;

QY      1 CTNCFQTPTPLWRRNPDPGLCNACGLFLKLHGVVRPLSLKTDVIKKNR 50
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      317 CANKCTTTTLLWRRNAGPVCNACGLYKLLHNVRPLTMKKEGIQTRNR 366

RESULT 15
AAU97801
ID      AAU97801 standard; Protein; 443 AA.
XX
AC      AAU97801;

```

Search completed: January 3, 2003, 19:47:49
Job time : 45.5 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January . 3, 2003, 15:22:02 ; Search time 2756 seconds
(without alignments)
15522.904 Million cell updates/sec

Title: US-09-725-010-1
Perfect score: 1470
Sequence: 1 gqatcttctcttttttcatttc.....cttttagtttcgaatatg 1470

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

total number of hits satisfying chosen parameters: 4109280

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*

1: gb_ba:*
2: gb_hug:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_man:*
37: em_htg_vrt:*
38: em_sy:*
39: em_hgtg_hum:*
40: em_hgtg_mus:*
41: em_hgtg_pln:*
42: em_hgtg_rod:*
43: em_hgtg_man:*
44: em_hgtg_vrt:*
45: em_hgtg_sy:*
46: em_hgtg_hum:*
47: em_hgtg_mus:*
48: em_hgtg_pln:*
49: em_hgtg_rod:*
50: em_hgtg_man:*
51: em_hgtg_vrt:*
52: em_hgtg_sy:*
53: em_hgtg_hum:*
54: em_hgtg_mus:*
55: em_hgtg_pln:*
56: em_hgtg_rod:*
57: em_hgtg_man:*
58: em_hgtg_vrt:*
59: em_hgtg_sy:*
60: em_hgtg_hum:*
61: em_hgtg_mus:*
62: em_hgtg_pln:*
63: em_hgtg_rod:*
64: em_hgtg_man:*
65: em_hgtg_vrt:*
66: em_hgtg_sy:*
67: em_hgtg_hum:*
68: em_hgtg_mus:*
69: em_hgtg_pln:*
70: em_hgtg_rod:*
71: em_hgtg_man:*
72: em_hgtg_vrt:*
73: em_hgtg_sy:*
74: em_hgtg_hum:*
75: em_hgtg_mus:*
76: em_hgtg_pln:*
77: em_hgtg_rod:*
78: em_hgtg_man:*
79: em_hgtg_vrt:*
80: em_hgtg_sy:*
81: em_hgtg_hum:*
82: em_hgtg_mus:*
83: em_hgtg_pln:*
84: em_hgtg_rod:*
85: em_hgtg_man:*
86: em_hgtg_vrt:*
87: em_hgtg_sy:*
88: em_hgtg_hum:*
89: em_hgtg_mus:*
90: em_hgtg_pln:*
91: em_hgtg_rod:*
92: em_hgtg_man:*
93: em_hgtg_vrt:*
94: em_hgtg_sy:*
95: em_hgtg_hum:*
96: em_hgtg_mus:*
97: em_hgtg_pln:*
98: em_hgtg_rod:*
99: em_hgtg_man:*
100: em_hgtg_vrt:*
101: em_hgtg_sy:*
102: em_hgtg_hum:*
103: em_hgtg_mus:*
104: em_hgtg_pln:*
105: em_hgtg_rod:*
106: em_hgtg_man:*
107: em_hgtg_vrt:*
108: em_hgtg_sy:*
109: em_hgtg_hum:*
110: em_hgtg_mus:*
111: em_hgtg_pln:*
112: em_hgtg_rod:*
113: em_hgtg_man:*
114: em_hgtg_vrt:*
115: em_hgtg_sy:*
116: em_hgtg_hum:*
117: em_hgtg_mus:*
118: em_hgtg_pln:*
119: em_hgtg_rod:*
120: em_hgtg_man:*
121: em_hgtg_vrt:*
122: em_hgtg_sy:*
123: em_hgtg_hum:*
124: em_hgtg_mus:*
125: em_hgtg_pln:*
126: em_hgtg_rod:*
127: em_hgtg_man:*
128: em_hgtg_vrt:*
129: em_hgtg_sy:*
130: em_hgtg_hum:*
131: em_hgtg_mus:*
132: em_hgtg_pln:*
133: em_hgtg_rod:*
134: em_hgtg_man:*
135: em_hgtg_vrt:*
136: em_hgtg_sy:*
137: em_hgtg_hum:*
138: em_hgtg_mus:*
139: em_hgtg_pln:*
140: em_hgtg_rod:*
141: em_hgtg_man:*
142: em_hgtg_vrt:*
143: em_hgtg_sy:*
144: em_hgtg_hum:*
145: em_hgtg_mus:*
146: em_hgtg_pln:*
147: em_hgtg_rod:*
148: em_hgtg_man:*
149: em_hgtg_vrt:*
150: em_hgtg_sy:*
151: em_hgtg_hum:*
152: em_hgtg_mus:*
153: em_hgtg_pln:*
154: em_hgtg_rod:*
155: em_hgtg_man:*
156: em_hgtg_vrt:*
157: em_hgtg_sy:*
158: em_hgtg_hum:*
159: em_hgtg_mus:*
160: em_hgtg_pln:*
161: em_hgtg_rod:*
162: em_hgtg_man:*
163: em_hgtg_vrt:*
164: em_hgtg_sy:*
165: em_hgtg_hum:*
166: em_hgtg_mus:*
167: em_hgtg_pln:*
168: em_hgtg_rod:*
169: em_hgtg_man:*
170: em_hgtg_vrt:*
171: em_hgtg_sy:*
172: em_hgtg_hum:*
173: em_hgtg_mus:*
174: em_hgtg_pln:*
175: em_hgtg_rod:*
176: em_hgtg_man:*
177: em_hgtg_vrt:*
178: em_hgtg_sy:*
179: em_hgtg_hum:*
180: em_hgtg_mus:*
181: em_hgtg_pln:*
182: em_hgtg_rod:*
183: em_hgtg_man:*
184: em_hgtg_vrt:*
185: em_hgtg_sy:*
186: em_hgtg_hum:*
187: em_hgtg_mus:*
188: em_hgtg_pln:*
189: em_hgtg_rod:*
190: em_hgtg_man:*
191: em_hgtg_vrt:*
192: em_hgtg_sy:*
193: em_hgtg_hum:*
194: em_hgtg_mus:*
195: em_hgtg_pln:*
196: em_hgtg_rod:*
197: em_hgtg_man:*
198: em_hgtg_vrt:*
199: em_hgtg_sy:*
200: em_hgtg_hum:*
201: em_hgtg_mus:*
202: em_hgtg_pln:*
203: em_hgtg_rod:*
204: em_hgtg_man:*
205: em_hgtg_vrt:*
206: em_hgtg_sy:*
207: em_hgtg_hum:*
208: em_hgtg_mus:*
209: em_hgtg_pln:*
210: em_hgtg_rod:*
211: em_hgtg_man:*
212: em_hgtg_vrt:*
213: em_hgtg_sy:*
214: em_hgtg_hum:*
215: em_hgtg_mus:*
216: em_hgtg_pln:*
217: em_hgtg_rod:*
218: em_hgtg_man:*
219: em_hgtg_vrt:*
220: em_hgtg_sy:*
221: em_hgtg_hum:*
222: em_hgtg_mus:*
223: em_hgtg_pln:*
224: em_hgtg_rod:*
225: em_hgtg_man:*
226: em_hgtg_vrt:*
227: em_hgtg_sy:*
228: em_hgtg_hum:*
229: em_hgtg_mus:*
230: em_hgtg_pln:*
231: em_hgtg_rod:*
232: em_hgtg_man:*
233: em_hgtg_vrt:*
234: em_hgtg_sy:*
235: em_hgtg_hum:*
236: em_hgtg_mus:*
237: em_hgtg_pln:*
238: em_hgtg_rod:*
239: em_hgtg_man:*
240: em_hgtg_vrt:*
241: em_hgtg_sy:*
242: em_hgtg_hum:*
243: em_hgtg_mus:*
244: em_hgtg_pln:*
245: em_hgtg_rod:*
246: em_hgtg_man:*
247: em_hgtg_vrt:*
248: em_hgtg_sy:*
249: em_hgtg_hum:*
250: em_hgtg_mus:*
251: em_hgtg_pln:*
252: em_hgtg_rod:*
253: em_hgtg_man:*
254: em_hgtg_vrt:*
255: em_hgtg_sy:*
256: em_hgtg_hum:*
257: em_hgtg_mus:*
258: em_hgtg_pln:*
259: em_hgtg_rod:*
260: em_hgtg_man:*
261: em_hgtg_vrt:*
262: em_hgtg_sy:*
263: em_hgtg_hum:*
264: em_hgtg_mus:*
265: em_hgtg_pln:*
266: em_hgtg_rod:*
267: em_hgtg_man:*
268: em_hgtg_vrt:*
269: em_hgtg_sy:*
270: em_hgtg_hum:*
271: em_hgtg_mus:*
272: em_hgtg_pln:*
273: em_hgtg_rod:*
274: em_hgtg_man:*
275: em_hgtg_vrt:*
276: em_hgtg_sy:*
277: em_hgtg_hum:*
278: em_hgtg_mus:*
279: em_hgtg_pln:*
280: em_hgtg_rod:*
281: em_hgtg_man:*
282: em_hgtg_vrt:*
283: em_hgtg_sy:*
284: em_hgtg_hum:*
285: em_hgtg_mus:*
286: em_hgtg_pln:*
287: em_hgtg_rod:*
288: em_hgtg_man:*
289: em_hgtg_vrt:*
290: em_hgtg_sy:*
291: em_hgtg_hum:*
292: em_hgtg_mus:*
293: em_hgtg_pln:*
294: em_hgtg_rod:*
295: em_hgtg_man:*
296: em_hgtg_vrt:*
297: em_hgtg_sy:*
298: em_hgtg_hum:*
299: em_hgtg_mus:*
300: em_hgtg_pln:*
301: em_hgtg_rod:*
302: em_hgtg_man:*
303: em_hgtg_vrt:*
304: em_hgtg_sy:*
305: em_hgtg_hum:*
306: em_hgtg_mus:*
307: em_hgtg_pln:*
308: em_hgtg_rod:*
309: em_hgtg_man:*
310: em_hgtg_vrt:*
311: em_hgtg_sy:*
312: em_hgtg_hum:*
313: em_hgtg_mus:*
314: em_hgtg_pln:*
315: em_hgt

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Query			ID	Description
			Match	Length	DB		
1	1470	100.0	1470	6	AX154641	Sequence	
2	505	34.4	2682	8	CAU64206	U64206 Candida alb	
3	153	10.4	2188	8	AF001978	Candida a	
C 4	136.2	9.3	42565	8	CAC35955	C. albican	
	91.2	6.2	86827	3	PFMAL3P5	AL034556 Plasmodiu	
C 5	86.8	5.9	104992	2	AC005504	Plasmodiu	
7	86.8	5.9	169546	2	AC004157	Plasmodiu	
C 8	75.2	5.1	104014	2	AC116921	Dictyoste	
	73.4	5.0	14867	3	AE001398	Plasmodiu	
10	73.4	5.0	164399	3	PFMAL3P6	Plasmodium	
11	72.8	5.0	113880	3	PFMAL3P4	Plasmodiu	
C 12	70.8	4.8	2009	6	AX457067	Sequence	
C 13	70.4	4.8	192929	2	AC005505	Plasmodiu	
C 14	69.4	4.7	56099	2	AC115598	Dictyoste	
C 15	69.2	4.7	253305	3	PFMAL3P7	Plasmodiu	
C 16	68.2	4.6	115489	2	AC117072	Dictyoste	
C 17	68	4.6	9810	6	AX345328	Sequence	
C 18	67.2	4.6	7347	1	AF211124	Carsonell	
C 19	67.2	4.6	67970	3	PFMAL1P3	AL031746 Plasmodiu	
C 20	67	4.6	130349	9	AC011593	Homo sapi	
C 21	67	4.6	198431	9	AC109594	Homo sapi	
C 22	66.8	4.5	97371	8	NT2888	Arabidops	
C 23	66.6	4.5	1141	6	AX083744	Sequence	
C 24	66.4	4.5	14635	3	AE001423	Plasmodiu	
C 25	66.4	4.5	18624	6	AX346604	Sequence	
C 26	66.4	4.5	30726	2	AC117269	Dictyoste	
27	66.2	4.5	1192	9	HS3323759	Homo sapi	
C 28	66.2	4.5	14867	3	AE001398	Plasmodiu	
C 29	66.2	4.5	113489	2	AC117072	Dictyoste	
C 30	66.2	4.5	133148	9	AC009435	Homo sapi	
C 31	66	4.5	17538	6	AX346059	Sequence	
C 32	66	4.5	201299	9	AC007158	Homo sapi	
C 33	65.8	4.5	156060	2	AC004153	Plasmodiu	
C 34	65.6	4.5	7584	6	AX286935	Sequence	
C 35	65.4	4.4	22977	9	AC106883	Homo sapi	
C 36	65.4	4.4	162515	9	AC107032	Homo sapi	
37	65.2	4.4	23676	2	AC115599	Dictyoste	
C 38	65.2	4.4	124820	2	AC117073	Dictyoste	
C 39	65.2	4.4	141017	2	AC116962	Dictyoste	
C 40	65.2	4.4	165260	9	AC024341	Homo sapi	
C 41	65	4.4	43993	2	AC116965	Dictyoste	
C 42	65	4.4	165059	9	AC106856	Homo sapi	
C 43	65	4.4	321003	2	PFMAL3P2	Plasmodiu	
C 44	64.8	4.4	1141	6	AX083744	Sequence	
C 45	64.8	4.4	5884	6	AX347067	Sequence	

ALIGNMENTS

RESULT 1					
AXI54641					
LOCUS		1470 bp	DNA	linear	PAT 23-JUN-2001
DEFINITION	Sequence 1 from Patent WO0138550.				
ACCESSION	AXI54641				
VERSION	AXI54641.1	GI:14536200			
KEYWORDS	.				
SOURCE	Candida albicans.				
ORGANISM	Candida albicans				
	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;				
	Saccharomycetales; mitosporic Saccharomycetales; Candida.				
REFERENCE	1 (bases 1 to 1470)				
AUTHORS	Sundstrom,P.				
TITLE	Methods for altering the expression of hyphal-specific genes				
JOURNAL	Patent: WO 0138550-A 1 31-MAY-2001.				

RESULT 4
CAC35A5/c

LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS

CAC35A5
C.albicans cosmid Ca35A5.
AL033396
AL033396.1 GI:3850143
ARS; Canikl; CDP-alcohol phosphatidyltransferase; chitin synthase;
chsi; cytochrome P450; DNA polymerase delta; histidine kinase;
isocitrate dehydrogenase pseudogene; LTR; pol3; Rcc1; regulator of
chromosome condensation; rehydrin; RPS1-like region.
Candida albicans.
Candida albicans

SOURCE
ORGANISM

Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; mitosporic Saccharomycetales; Candida.
1 (bases 1 to 42565)
Tait, E., Simon, M.C., King, S., Brown, A.J., Gow, N.A. and Shaw, D.J.

REFERENCE
AUTHORS

A Candida albicans genome project: cosmid contigs, physical
mapping, and gene isolation

JOURNAL
MEDLINE

Fungal Genet. Biol. 21 (3), 308-314 (1997)
97435544

REMARK
PUBMED

9290243
Article No. FG970983

REFERENCE
AUTHORS

2 (bases 1 to 42565)
Oliver, K. and Harris, D.

JOURNAL
AUTHORS

Unpublished
3 (bases 1 to 42565)
Barrell, B.G. and Rajandream, M.A.

JOURNAL
TITLE

Submitted (05-NOV-1998) On behalf of the pilot sequencing project
on the Candida albicans strain 1161 genome. Sanger Centre, Wellcome
Trust Genome Campus, Hinxton, Cambridge CB10 1SA E-mail:
barrell@sanger.ac.uk Cosmids supplied by Prof. Duncan Shaw, [3]
Department of Molecular and Cell Biology, The Institute of Medical
Science, University of Aberdeen, Foresterhill, Aberdeen, AB25 2ZD,
UK

COMMENT

Notes:
Funding: sequencing funded by Beowulf Genomics Ltd. CDS are
numbered using the following system eg CAC20C1.01c. CA (C.
albicans), 20C1 (cosmid name),
.01 (first CDS), c (complementary strand).
The more significant matches with motifs in the PROSITE/PRAM
database are also included but some of these may be fortuitous.
IMPORTANT: This sequence MAY NOT be the entire insert of the
sequenced clone. It may be shorter because we only sequence
overlapping sections once, or longer, because we arrange for a
small overlap between neighbouring submissions. Cosmid Ca35A5 is
likely to map to region R, chromosome 7.

FEATURES
source

1..42565
/organism="Candida albicans"
/strain="1161"
/db_xref="taxon:5476"
/map="chromosome 7"
/clone="cosmid Ca35A5"
complement(1..785)
/gene="Ca35A5.01c"
complement(<1..785)
/gene="Ca35A5.01c"
/note="Ca35A5.01c, partial orf, len: > 261 aa, most
similar to EPT1.YEAST ethanolaminephosphotransferase (EC
2.7.8.1) (391 aa), fasta scores opt: 984, E(): 0, (56.3%
identity in 263 aa overlap), contains PS00379 CDP-alcohol
phosphatidyltransferases signature"
/codon_start=1
/transl_table=12
/product="putative alcohol phosphatidyl transferase"
/protein_id="CAA21944.1"
/db_xref="GI:3850144"
/translation="MGLEFPTNKLONLKYKSSDEHSIIISYILKKWNFVQIFPL
SMAPNVYLLGLFFIIGNLWTFYIDYLNSTQPTWCYFFAGFLMFYQTFDGCDCGH
ARRTQSGELFDHSDAINTLTGTFVFASVLKMGYGLLLLSQFASVCNFTYSTW
EEXYHTLFLSKSGPVEGILMVICIWIITGIFGPDITIDLFELNLTSLGYGYKVD

misc_feature

/complement(423..491)
/gene="Ca35A5.01c"
/note="PS00379 CDP-alcohol phosphatidyltransferases
signature"
4307..11124

misc_feature

/note="RPS1-like region"

misc_feature

complement(7606..8085)

misc_feature

/note="region showing similarity with isocitrate
dehydrogenase"

gene

13175..14713

CDS

/gene="Ca35A5.02"
13175..14713

CDS

/note="Ca35A5.02"
13175..14713

CDS

/note="Ca35A5.02, unknown, len: 512 aa, possible
glycoprotein"

CDS

/codon_start=1
/transl_table=12

CDS

/product="hypothetical protein"
/protein_id="CAA21945.1"

CDS

/db_xref="GI:3850145"

CDS

/translation="MSEVFNQSPSSILKQSNASSSTKVHGAPSSQIANVDFTNLR
KLSNTFSNLTNTNTNSALSRFLTRNKSLSNISVHPSSDDSDSPKLTRESTSP
SEKSVSGNRLRIAKLFKSKNSQSRKPDLDFTSSITSDSSFRKVIIGNSLNE
TKPRKSSPMSTTFHSLFRSHSHGSLQDRDQVATGTTPLGKDFDFKSAKTT
LCLSSNSNIIISNPDLAIOYNTNPNISIEDRETLDHTNSSFIDHKKMLVPADSF
IONKLNKYHOTEGVLGIVSELDHNDANKIYNLYLKLPLTFPSISDSQSKSMR
PILSAVERIANFVKEFCLHQPNERFSRKTSSVSLGRKDFEDFDLQLSLFE
KMLSLSHLQITFESEVSLQALILNWKYINAYRVYLLLSIFQLIYLNLEFTRSH
NSKIIIRDLLLVSRFKVFTTEQIGSGERETSQFLGNAESEDITLGNLTSTLAVL
SSIS"

gene

18031..18456

CDS

/gene="Ca35A5.03"
18031..18456

CDS

/note="Ca35A5.03, unknown, questionable orf, len: 141 aa"

CDS

/codon_start=1
/transl_table=12

CDS

/product="questionable orf"
/protein_id="CAA21946.1"

CDS

/db_xref="GI:3850146"

CDS

/translation="MNQNNKSNTERNSIEEDALHYHYEYSVSLVOHEKKNQOEIS
MVEIRLRKANSEYRDSIDIKELIVNDQKDSQNEAFSSLMKLSALKQVDKNYOLKI
ENKRLKELVDAQARLVDFKDEWICFGPELKKIRKQKV"
19043..19210

misc_feature

/note="BLASTN match (score: 669, strand:(+1), range:
8-175) to AF069450 C. albicans retrotransposon long
terminal repeat zeta"

misc_feature

19204..19277
/note="BLASTN hit (score: 325, strand:(+1), range:
168-24) to AF069450 C. albicans retrotransposon long
terminal repeat zeta"

misc_feature

20359..20769
/note="BLASTN match (score: 1704 strand:(+1), range:
49-459) to CAARESE X65035 C.albicans DNA of an
autonomously replicating sequence (ARS)"

misc_feature

20719..21001
/note="BLASTN match (score: 1298 strand:(+1), range:
405-687) to CAARESE X65035 C.albicans DNA of an
autonomously replicating sequence (ARS)"

gene

20808..23888

CDS

/gene="Chs1"
20808..23888

CDS

/note="Chs1"
/gene="Ca35A5.04, Chs1 gene, len: 1026 aa, CHS1_CANAL
chitin synthase 1 (EC 2.4.1.16), CHS1_CANAL (776 aa)
(99.9% identity in 776 aa overlap) but open reading frame
here extends 250 aa upstream of CHS1_CANAL start
at 21561"

CDS

/codon_start=1
/transl_table=12
/product="chitin synthase"
/protein_id="CAA21947.1"

GenCore version 5.1.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 3, 2003, 19:47:58 ; Search time 15.5 Seconds
(without alignments)
94.913 Million cell updates/sec

Title: US-09-725-010-3

Perfect score: 281

Sequence: 1 CTNCTQTTPPLWRNPDPGP.....LHGVRPLSLKTDVIKKNR 50

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep.*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	277	98.6	853	3	US-09-254-325-2
2	277	98.6	856	3	US-09-079-415-6
3	277	98.6	866	3	US-08-750-458A-2
4	188	66.9	449	2	US-08-927-394-2
5	53	18.9	108	1	US-08-204-740-11
6	53	18.9	108	3	US-09-081-167A-11
7	53	18.9	108	3	US-09-081-395-11
8	53	18.9	108	4	US-09-416-833-11
9	53	18.9	108	5	PCT-US95-02521-11
10	52.5	18.7	2237	1	US-08-455-543A-48
11	52.5	18.7	2237	2	US-08-223-305C-48
12	52.5	18.7	2237	4	US-09-268-163-8
13	52	18.5	128	5	PCT-US95-06266-141
14	49	17.4	314	2	US-08-460-309-19
15	49	17.4	314	2	US-08-125-077-19
16	49	17.4	664	2	US-08-852-153-8
17	49	17.4	786	4	US-09-103-429A-3
18	49	17.4	805	4	US-09-103-429A-4
19	49	17.4	1068	3	US-08-390-874C-11
20	49	17.4	1068	4	US-09-265-772-11
21	49	17.4	1069	2	US-08-162-081B-37
22	49	17.4	1069	2	US-08-780-872-37
23	49	17.4	1069	4	US-09-085-957-37
24	49	17.4	1080	2	US-08-162-081B-36
25	49	17.4	1080	2	US-08-780-872-36
26	49	17.4	1080	4	US-09-085-957-36
27	48.5	17.3	63	1	US-08-269-441A-13

28	48.5	17.3	314	1	US-08-269-441A-2	Sequence 2, Appli
29	48.5	17.3	314	4	US-09-008-465-3	Sequence 3, Appli
30	48.5	17.3	314	4	US-09-276-851-2	Sequence 2, Appli
31	48.5	17.3	314	4	US-09-528-939-3	Sequence 3, Appli
32	48.5	17.3	785	3	US-08-374-077C-3	Sequence 3, Appli
33	48.5	17.3	785	4	US-08-895-590-3	Sequence 3, Appli
34	48.5	17.3	785	4	US-09-539-879A-3	Sequence 2, Appli
35	48.5	17.3	2516	3	US-08-374-077C-2	Sequence 3, Appli
36	48.5	17.3	2516	4	US-08-895-590-2	Sequence 2, Appli
37	48.5	17.3	2516	4	US-09-539-879A-2	Sequence 2, Appli
38	48.5	17.3	2972	4	US-09-579-181-2	Sequence 2, Appli
39	48.5	17.3	3118	4	US-09-579-181-1	Sequence 2, Appli
40	48	17.1	256	2	US-08-484-905-114	Sequence 114, App
41	48	17.1	256	3	US-08-481-985B-114	Sequence 114, App
42	48	17.1	256	4	US-08-370-476-114	Sequence 114, App
43	48	17.1	431	3	US-08-478-507-2	Sequence 2, Appli
44	48	17.1	431	4	US-09-128-275A-2	Sequence 2, Appli
45	48	17.1	431	4	US-09-553-427-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-09-254-325-2
; Sequence 2, Application US/09254325
; Patent No. 6090607
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: ENHANCED EXPRESSION OF
; TITLE OF INVENTION: PROTEOLYTIC ENZYMES IN KOJI MOLDS
; NUMBER OF SEQUENCES: 17
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/254,325
; FILING DATE:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 853 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FEATURE:
; NAME/KEY: Binding-site
; LOCATION: 652-676
; OTHER INFORMATION: /note= "DNA BINDING SITE"
; FEATURE:
; NAME/KEY: Region
; LOCATION: 1..731
; OTHER INFORMATION: /note= "TRUNCATED AREA WHICH IS
; OTHER INFORMATION: STILL ACTIVE BUT NOT REPRESSED BY
; OTHER INFORMATION: L-GLUTAM..."
US-09-254-325-2

Query Match 98.6%; Score 277; DB 3; Length 853;
Best Local Similarity 98.0%; Pred. No. 1.3e-30;
Matches 49; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTNCTQTTPPLWRNPDPGP...LHGVRPLSLKTDVIKKNR 50
|||||
Db 652 CTNCTQTTPPLWRNPDPGP...LHGVRPLSLKTDVIKKNR 701

RESULT 2
US-09-079-415-6
; Sequence 6, Application US/09079415
; Patent No. 6013452
; GENERAL INFORMATION:

APPLICANT: Christensen, Tove
APPLICANT: Lehmebeck, Jan
TITLE OF INVENTION: A Fungus Wherein The area, pepC and/or
TITLE OF INVENTION: pepE Genes Have Been Inactivated
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 60134520 No. 6013452disk of No. 6013452th America, Inc.
STREET: 405 Lexington Avenue
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10174
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/079,415
FILING DATE: 14-MAY-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Rozek, Carol
REGISTRATION NUMBER: 36,993
REFERENCE/DOCKET NUMBER: 4657-204-US
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 866 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-079-415-6

Query Match 98.6%; Score 277; DB 3; Length 866;
Best Local Similarity 98.0%; Pred. No. 1.4e-30;
Matches 49; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTNCFQTPTPLWRNPDPGLCNACGLFLKLGHVVRPLSLKTDVTKRNR 50
|||||
DB 664 CTNCFQTPTPLWRNPDPGLCNACGLFLKLGHVVRPLSLKTDVTKRNR 713

RESULT 3
US-08-750-458A-2
Sequence 2, Application US/08750458A
Patent No. 6025185
GENERAL INFORMATION:
APPLICANT: Christensen, Tove
APPLICANT: Hynes, Michael J.
TITLE OF INVENTION: A Fungus Wherein The Area Gene Has Been Modified And An Area
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 60251850 No. 6025185disk of No. 6025185th America, Inc.
STREET: 405 Lexington Avenue, Suite 6400
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10174-6401
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/750,458A
FILING DATE: 3-December-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Rozek, Carol E.
REGISTRATION NUMBER: 36,993

REFERENCE/DOCKET NUMBER: 4129-204-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212 867 0123
TELEFAX: 212 867 0298
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 866 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-750-458A-2

Query Match 98.6%; Score 277; DB 3; Length 866;
Best Local Similarity 98.0%; Pred. No. 1.4e-30;
Matches 49; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTNCFQTPTPLWRNPDPGLCNACGLFLKLGHVVRPLSLKTDVTKRNR 50
|||||
DB 664 CTNCFQTPTPLWRNPDPGLCNACGLFLKLGHVVRPLSLKTDVTKRNR 713

RESULT 4
US-08-927-394-2
Sequence 2, Application US/08927394
Patent No. 5990092
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: GATA-6 TRANSCRIPTION FACTOR: COMPOSITIONS AND METHODS
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Wolf, Greenfield, & Sacks, P.C.
STREET: 600 Atlantic Avenue
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02210
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/927,394
FILING DATE:
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/025,574
FILING DATE: 06-SEP-1996
ATTORNEY/AGENT INFORMATION:
NAME: Plumer, Elizabeth R.
REGISTRATION NUMBER: 36,637
REFERENCE/DOCKET NUMBER: S1237/7005
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-720-3500
TELEFAX: 617-720-2441
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 449 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-927-394-2

Query Match 66.9%; Score 188; DB 2; Length 449;
Best Local Similarity 62.0%; Pred. No. 2.1e-18;
Matches 31; Conservative 9; Mismatches 10; Indels 0; Gaps 0;

QY 1 CTNCFQTPTPLWRNPDPGLCNACGLFLKLGHVVRPLSLKTDVTKRNR 50
|||||
DB 298 CANCHTTTTLWRNAGEPVCNACGLYMKLHGVPRLAMKKEGIQTRKR 347

RESULT 5

US-08-204-740-11
; Sequence 11, Application US/08204740
; Patent No. 5753432
; GENERAL INFORMATION:
; APPLICANT: Gudkov, Andrei
; APPLICANT: Kazarov, Alexander
; APPLICANT: Mazo, Ilya
; APPLICANT: Roninson, Igor B
; TITLE OF INVENTION: Methods for Identifying Genetic
; TITLE OF INVENTION: Suppressor Elements and Genes Associated with Malignant
; TITLE OF INVENTION: Growth in Cancer Cells
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Allgretti & Witcoff, Ltd.
; STREET: 10 S. Wacker Drive, Suite 3000
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/204,740
; FILING DATE: 04-MAR-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 5753432nan, Kevin E
; REGISTRATION NUMBER: 35,303
; REFERENCE/DOCKET NUMBER: 93,354-C
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-715-1000
; TELEFAX: 312-715-1234
; TELEX: 910-221-5317
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 108 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-204-740-11

Query Match 18.9%; Score 53; DB 1; Length 108;
Best Local Similarity 43.5%; Pred. No. 3.8;
Matches 10; Conservative 7; Mismatches 6; Indels 0; Gaps 0;

Qy 28 FLKLHGVRPLSLKTDVKKRR 50
||: : ||::||: : || |
Db 14 FLEANEVPRVTLRTNLTTRR 36

RESULT 6
US-09-081-167A-11
; Sequence 11, Application US/09081167A
; Patent No. 6083745
; GENERAL INFORMATION:
; APPLICANT: Gudkov, Andrei
; APPLICANT: Kazarov, Alexander
; APPLICANT: Mazo, Ilya
; APPLICANT: Roninson, Igor B
; TITLE OF INVENTION: Methods for Identifying Genetic
; TITLE OF INVENTION: Suppressor Elements and Genes Associated with Malignant
; TITLE OF INVENTION: Growth in Cancer Cells
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff
; STREET: 300 S. Wacker Drive, 32nd Floor
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/081,167A
; FILING DATE: 18-MAY-1998
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 6083745nan, Kevin E
; REGISTRATION NUMBER: 35,303
; REFERENCE/DOCKET NUMBER: 93,354-KK
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-913-0001
; TELEFAX: 312-913-0002
; TELEX:
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 108 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-09-081-167A-11

Query Match 18.9%; Score 53; DB 3; Length 108;
Best Local Similarity 43.5%; Pred. No. 3.8;
Matches 10; Conservative 7; Mismatches 6; Indels 0; Gaps 0;

Qy 28 FLKLHGVRPLSLKTDVKKRR 50
||: : ||::||: : || |
Db 14 FLEANEVPRVTLRTNLTTRR 36

RESULT 7
US-09-081-395-11
; Sequence 11, Application US/09081395
; Patent No. 6083746
; GENERAL INFORMATION:
; APPLICANT: Gudkov, Andrei
; APPLICANT: Kazarov, Alexander
; APPLICANT: Mazo, Ilya
; APPLICANT: Roninson, Igor B
; TITLE OF INVENTION: Methods for Identifying Genetic
; TITLE OF INVENTION: Suppressor Elements and Genes Associated with Malignant
; TITLE OF INVENTION: Growth in Cancer Cells
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff
; STREET: 300 S. Wacker Drive, 32nd Floor
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/081,395
; FILING DATE: 18-MAY-1998
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 6083746nan, Kevin E
; REGISTRATION NUMBER: 35,303
; REFERENCE/DOCKET NUMBER: 93,354-KK
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-913-0001
; TELEFAX: 312-913-0002
; TELEX:
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:

; LENGTH: 108 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-09-081-395-11

Query Match 18.9%; Score 53; DB 3; Length 108;
Best Local Similarity 43.5%; Pred. No. 3.8;
Matches 10; Conservative 7; Mismatches 6; Indels 0; Gaps 0;

QY 28 FLKLHGVVRPLSLKTDVTKRNR 50
||: : | ||: ||: : | || |
Db 14 FLEANEVPRPVTLRTNLTNR 36

RESULT 8

US-09-416-833-11
; Sequence 11, Application US/09416833
; Patent No. 6197521

; GENERAL INFORMATION:
; APPLICANT: Gudkov, Andrei

; APPLICANT: Kazarov, Alexander
; APPLICANT: Mazo, Iliya

; APPLICANT: Roninson, Igor B
; TITLE OF INVENTION: Methods for Identifying Genetic

; TITLE OF INVENTION: Suppressor Elements and Genes Associated with Malignant

; TITLE OF INVENTION: Growth in Cancer Cells
; NUMBER OF SEQUENCES: 13

; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Alligretti & Witcoff, Ltd.

; STREET: 10 S. Wacker Drive, Suite 3000
; CITY: Chicago

; STATE: Illinois
; COUNTRY: USA

; ZIP: 60606
; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/416,833

; FILING DATE:
; CLASSIFICATION:

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/204,740

; FILING DATE:
; ATTORNEY/AGENT INFORMATION:

; NAME: No. 6197521nan, Kevin E
; REGISTRATION NUMBER: 35,303

; REFERENCE/DOCKET NUMBER: 93,354-C
; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 312-715-1000
; TELEFAX: 312-715-1234

; TELEX: 910-221-5317
; INFORMATION FOR SEQ ID NO: 11:

; SEQUENCE CHARACTERISTICS:
; LENGTH: 108 amino acids

; TYPE: amino acid
; TOPOLOGY: linear

; MOLECULE TYPE: peptide
US-09-416-833-11

Query Match 18.9%; Score 53; DB 4; Length 108;
Best Local Similarity 43.5%; Pred. No. 3.8;
Matches 10; Conservative 7; Mismatches 6; Indels 0; Gaps 0;

QY 28 FLKLHGVVRPLSLKTDVTKRNR 50
||: : | ||: ||: : | || |
Db 14 FLEANEVPRPVTLRTNLTNR 36

RESULT 9

PCT-US95-02521-11
; Sequence 11, Application PC/TUS9502521
; GENERAL INFORMATION:

; APPLICANT:

; TITLE OF INVENTION: Methods for Identifying Genetic

; TITLE OF INVENTION: Suppressor Elements and Genes Associated with Malignant

; NUMBER OF SEQUENCES: 13

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent in Release #1.0, Version #1.25 (EPO)

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: PCT/US95/02521

; FILING DATE:

; INFORMATION FOR SEQ ID NO: 11:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 108 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: peptide

PCT-US95-02521-11

Query Match 18.9%; Score 53; DB 5; Length 108;
Best Local Similarity 43.5%; Pred. No. 3.8;
Matches 10; Conservative 7; Mismatches 6; Indels 0; Gaps 0;

QY 28 FLKLHGVVRPLSLKTDVTKRNR 50
||: : | ||: ||: : | || |

Db 14 FLEANEVPRPVTLRTNLTNR 36

RESULT 10

US-08-455-543A-48

; Sequence 48, Application US/08455543A
; Patent No. 5792846

; GENERAL INFORMATION:

; APPLICANT: Harpold, Michael

; APPLICANT: Ellis, Steven

; APPLICANT: Williams, Mark

; APPLICANT: Feldman, Daniel

; APPLICANT: McCue, Ann

; APPLICANT: Brenner, Robert

; TITLE OF INVENTION: HUMAN CALCIUM CHANNEL COMPOSITIONS AND

; NUMBER OF SEQUENCES: 57

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Brown, Martin, Haller & McClain

; STREET: 1660 Union Street

; CITY: San Diego

; STATE: California

; COUNTRY: USA

; ZIP: 92101-2926

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: DOS

; SOFTWARE: FastSeq Version 1.5

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/455,543A

; FILING DATE: May 31, 1995

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/223,305

; FILING DATE: April 4, 1994

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 07/868,354

; FILING DATE: April 10, 1992

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/745,206

; FILING DATE: 15-AUG-1991

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/620,250


```

> GENERAL INFORMATION:
> APPLICANT: Engvall, Eva
> APPLICANT: Leivo, Ilmo
> TITLE OF INVENTION: Nucleic Acids Encoding Merosin, Merosin
> TITLE OF INVENTION: Fragments and Uses Thereof
> NUMBER OF SEQUENCES: 23
> CORRESPONDENCE ADDRESS:
> ADDRESSEE: Campbell and Flores
> STREET: 4370 La Jolla Village Drive, Suite 700
> CITY: San Diego
> STATE: California
> COUNTRY: USA
> ZIP: 92122
> COMPUTER READABLE FORM:
> MEDIUM TYPE: Floppy disk
> COMPUTER: IBM PC compatible
> OPERATING SYSTEM: PC-DOS/MS-DOS
> SOFTWARE: PatentIn Release #1.0, Version #1.25
> CURRENT APPLICATION DATA:
> APPLICATION NUMBER: US/08/460,309
> FILING DATE:
> CLASSIFICATION: 435
> PRIORITY INFORMATION DATA:
> APPLICATION NUMBER: US 08/125,077
> FILING DATE: 22-SEP-1993
> APPLICATION NUMBER: US PCT/US 94/10730
> FILING DATE: 21-SEP-1994
> PRIORITY INFORMATION DATA:
> APPLICATION NUMBER: US 07/472,319
> FILING DATE: 30-JAN-1990
> PRIORITY INFORMATION DATA:
> APPLICATION NUMBER: US 07/919,951
> FILING DATE: 27-JUL-1992
> ATTORNEY/AGENT INFORMATION:
> NAME: Campbell, Cathryn A.
> REGISTRATION NUMBER: 31,815
> REFERENCE/DOCKET NUMBER: P-LA 9721
> TELECOMMUNICATION INFORMATION:
> TELEPHONE: (619) 535-9001
> TELEFAX: (619) 535-8949
> INFORMATION FOR SEQ ID NO: 19:
> SEQUENCE CHARACTERISTICS:
> LENGTH: 314 amino acids
> TYPE: amino acid
> TOPOLOGY: linear
> US-08-460-309-19

Query Match 17.4%; Score 49; DB 2; Length 314;
Best Local Similarity 27.3%; Pred. No. 50;
Matches 15; Conservative 4; Mismatches 12; Indels 24;

QY 1 CTNC---FTQTTLWRNPQGQL-CNAC-----GULFLKLG 33
  | | | | | | | | | | | | | | | | | | | | | |
Db 35 CNECCPGFEQK--WRQNTNARDFNCEPCNCHGHSNECKYDEVNRKGLSLDING 87
  | | | | | | | | | | | | | | | | | | | | | |

RESULT 15
US-08-125-077-19
> Sequence 19, Application US/08125077
> Patent No. 5872231
> Patent No. 5872231 5840863
> GENERAL INFORMATION:
> APPLICANT: Engvall, Eva
> APPLICANT: Leivo, Ilmo
> TITLE OF INVENTION: Nucleic Acids Encoding Merosin, Merosin
> TITLE OF INVENTION: Fragments and Uses Thereof
> NUMBER OF SEQUENCES: 23
> CORRESPONDENCE ADDRESS:
> ADDRESSEE: Campbell and Flores
> STREET: 4370 La Jolla Village Drive, Suite 700
> CITY: San Diego
> STATE: California
> COUNTRY: USA

```

```

Query Match      18.5%; Score 52; DB 5; Length 128;
Best Local Similarity 31.2%; Pred. No. 6.5;
Matches 15; Conservative 10; Mismatches 21; Indels 2; Gaps 1;

QY      1  CTNCGTQTPTPLWRRNPDCQPLNCAGLFLKHLGVVRPLSKLTDVIKKR 48
      |  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db      37  CMMPSTRTPTLYPRWQDTKGIPSCGMSLAFSQVLK--SLNTSHIQSQ 82

RESULT 14
US-08-460-309-19
; Sequence 19, Application US/08460309
; Patent No. 5837496

```

ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/125,077
FILING DATE: 22-SEP-1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US PCT/US 94/10730
FILING DATE: 21-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/472,319
FILING DATE: 30-JAN-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/919,951
FILING DATE: 27-JUL-1992
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-LA 9721
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 19:
SEQUENCE CHARACTERISTICS:
LENGTH: 314 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-125-077-19

Query Match 17.4%; Score 49; DB 2; Length 314;
Best Local Similarity 27.3%; Pred. NO. 50;
Matches 15; Conservative 4; Mismatches 12; Indels 24; Gaps 4;
QY 1 CTNC---FTQTPLWRNPDPGQPL-CNAC-----GLFLKLHG 33
Db 35 CNECCPGFEQKK--WRQNTNARPNCEPCNCHGHSNECKYDEEVNKGSLDING 87

Search completed: January 3, 2003, 19:52:38
Job time : 16.5 secs

A:Reference number: Z21970
A:Accession: T41336
A>Status: translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-170 <LYN>
A:Cross-references: EMBL:AL035076; NID:g4107277; PIDN:CAA22647.1; PID:g4107278; GSPDPB
A:Experimental source: strain 972n.; cosmid c417
R:Seeger, K.; Harris, D.; Wood, V.; Rajandream, M.A.; Barrell, B.G.
submitted to the EMBL Data Library, March 1999
A:Reference number: Z21928
A:Accession: T41208
A:Molecule type: DNA
A:Residues: 129-855 <SEE>
A:Cross-references: EMBL:AL049521; NTD:g6318249; PIDN:CAB40003.1; PID:g4539591; GSPDB
R:Hoe, K.L.; Won, M.S.; Chung, K.S.; Park, S.K.; Kim, D.U.; Jang, Y.J.; Yoo, O.J.; YO
Gene 215, 319-328, 1998
A>Title: Molecular cloning of gafI, a Schizosaccharomyces pombe GATA factor, which ca
A:Reference number: Z22401; MUID:98382525; PMID:9714631
A:Accession: T43297
A>Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 566-855 <HOB>
A:Cross-references: EMBL:L31601; NID:g710606; PIDN:AAC35593.1; PID:g710607
C:Genetics:
A:Gene: SPDB:SPCC417.01c; SPDB:SPCC1902.01; gafI
A:Map position: 3R
C:Superfamily: GATA-type zinc finger homology
C:Keywords: transcription factor; transcription regulation; zinc finger
F:632-685/Domain: GATA-type zinc finger homology <GZF>

Query Match 86.7%; Score 242; DB 2; Length 855;
Best Local Similarity 84.0%; Pred. No. 6,3e-22;
Matches 42; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

QY 1 CTNCGTKTTPLWRNPQGOLPCNACGLFLKLHGCVVRPLSLKTVDVIKKRR 50
|||||:||||||:||||||:||||||:||||||:||||||:||||||:||||||
Db 635 CTCNQCTTTLWRNSPDGOLPCNACGLFMKINGVVRPLSLKTVDIKKNR 684

RESULT 5
S56233
probable membrane protein YFL021w - yeast (Saccharomyces cerevisiae)
C:Species: Saccharomyces cerevisiae
C>Date: 02-Sep-1995 #sequence_revision 19-Oct-1995 #text_change 06-Feb-1998
C:Accession: S56233
R:Murakami, Y.; Naitou, M.; Hagiwara, H.; Shibata, T.; Ozawa, M.; Sasanuma, S.I.; Sas
submitted to the EMBL Data Library, May 1995
A:Description: Analysis of the nucleotide sequence of chromosome VI from Saccharomyces
A:Reference number: S56186
A:Accession: S56233
A:Molecule type: DNA
A:Residues: 1-510 <MUR>
A:Cross-references: EMBL:D50617; NID:g836685; PID:d1009858; PID:g836733; MIPS:YFL021w
C:Genetics:
A:Gene: SGD:GAT1
A:Cross-references: SGD:S0001873; MIPS:YFL021w
A:Map position: 6L
C:Superfamily: GATA-type zinc finger homology
C:Keywords: transmembrane protein; zinc finger
F:14-30/Domain: transmembrane #status predicted <TM1>
F:307-360/Domain: GATA-type zinc finger homology <GZF>

Query Match 85.3%; Score 238; DB 2; Length 510;
Best Local Similarity 82.0%; Pred. No. 1.2e-21;
Matches 41; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 1 CTNCGTKTTPLWRNPQGOLPCNACGLFLKLHGCVVRPLSLKTVDVIKKRR 50
:|||:|||||:|:|||||:|||||:|||||:|||||:|||||:|||||
Db 310 CSNCTSTTTLWRNKDQGLPLCNACGLFLKLHGCVVRPLSLKTVDIIKKRR 359

RESULT 6
S50543

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 3, 2003, 19:42:57 ; Search time 48 seconds
(without alignments)
214.632 Million cell updates/sec

Title: US-09-725-010-3

Perfect score: 281

Sequence: 1 CTNCFQTTPPLWRNPDGQP.....LHGVRPLSLKTDVIKRRN 50

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_21:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phage:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_virus:*
- 16: sp_bacteriap:*
- 17: sp_archaea:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	277	98.6	853	3 Q9C414	Q9C414 aspergillus
2	277	98.6	866	3 Q9Y7E8	Q9Y7E8 aspergillus
3	277	98.6	918	3 Q9HEW7	Q9HEW7 cladosporiu
4	277	98.6	944	3 Q60043	Q60043 metarhizium
5	247	87.9	1075	3 Q9P952	Q9P952 schizoshyll
6	240	85.4	327	3 P87016	P87016 zygosacchar
7	204	72.6	298	3 Q59842	Q59842 penicillium
8	198	70.5	312	3 Q9HEV2	Q9HEV2 emericella
9	198	70.5	320	3 Q9HEV3	Q9HEV3 emericella
10	198	70.5	436	3 Q9HEV4	Q9HEV4 emericella
11	195	69.4	383	13 Q9W6U0	Q9W6U0 brachydanio
12	194	69.0	119	5 Q77157	Q77157 strongyloce
13	194	69.0	537	5 Q95VY5	Q95VY5 anopheles g
14	193	68.7	486	5 Q8T3J0	Q8T3J0 drosophila
15	190	67.6	324	5 Q9NJ18	Q9NJ18 ceratitis c
16	190	67.6	511	5 Q8WQ4	Q8WQ4 calliphora

17	190	67.6	949	5 Q94884	Q94884 drosophila
18	190	67.6	950	5 Q9VF01	Q9VF01 drosophila
19	188	66.9	418	13 Q90410	Q90410 brachydanio
20	188	66.9	446	6 Q95JA5	Q95JA5 sus scrofa
21	185	65.8	338	13 Q9PTJ2	Q9PTJ2 brachydanio
22	184	65.5	383	13 Q9PTJ1	Q9PTJ1 brachydanio
23	183	65.1	266	6 Q95KV6	Q95KV6 bos taurus
24	183	65.1	431	5 Q77156	Q77156 strongyloce
25	183	65.1	439	4 Q16365	Q16365 homo sapien
26	183	65.1	442	11 Q8VI87	Q8VI87 mus musculu
27	182	64.8	116	6 Q9MYT2	Q9MYT2 oryctolagus
28	179	63.7	118	11 Q9QW52	Q9QW52 mus sp. ret
29	179	63.7	480	4 Q96BH0	Q96BH0 homo sapien
30	179	63.7	480	11 Q9DC59	Q9DC59 mus musculu
31	179	63.7	480	11 Q9DBY9	Q9DBY9 mus musculu
32	179	63.7	480	11 Q924Y4	Q924Y4 rattus norv
33	179	63.7	480	11 Q8VHY4	Q8VHY4 rattus norv
34	178	63.3	194	13 Q90411	Q90411 brachydanio
35	178	63.3	444	11 Q99NH5	Q99NH5 rattus norv
36	178	63.3	444	13 Q8QFW5	Q8QFW5 raja. eglant
37	178	63.3	549	3 Q9Y754	Q9Y754 emericella
38	176	62.6	704	5 Q9BPQ3	Q9BPQ3 drosophila
39	176	62.6	734	5 Q9VF00	Q9VF00 drosophila
40	175	62.3	388	3 Q9C154	Q9C154 botrytis ci
41	173	61.6	868	5 Q9NFF4	Q9NFF4 aedes aegyp
42	171	60.9	587	3 Q74262	Q74262 neurospora
43	165	58.7	232	5 Q8SUM6	Q8SUM6 encephalito
44	160	56.9	352	5 Q9BPQ1	Q9BPQ1 drosophila
45	158	56.2	466	4 Q96BH8	Q96BH8 homo sapien

ALIGNMENTS

RESULT 1

Q9C414 PRELIMINARY; PRT; 853 AA.
 ID Q9C414;
 AC Q9C414;
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE Area.
 OS AREA.
 GN Aspergillus oryzae.
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
 OC Eukaryotes; Trichocomaceae; mitosporic Trichocomaceae; Aspergillu.
 OX NCBI_TaxID=5062;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=TK3;
 RA van den Broek P.;
 RT "The A. oryzae area gene as a tool to overproduce a large number of
 proteases.";
 RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF320305; AAK08066.1; -;
 DR HSSP; P17429; 4GAT.
 DR InterPro; IPR000679; Znf_GATA.
 DR Pfam; PF00320; GATA; 1.
 DR PRINTS; PR00619; GATAZNFINGER.
 DR SMART; SM00401; Znf_GATA; 1.
 DR PROSITE; PS00344; GATA_ZN_FINGER_1; 1.
 DR PROSITE; PS01114; GATA_ZN_FINGER_2; 1.
 SQ SEQUENCE 853 AA; 91497 MW; A5EC8ICE256A6B15 CRC64;

Query Match 98.6%; Score 277; DB 3; Length 853;
 Best Local Similarity 98.0%; Pred. No. 8.6e-30;
 Matches 49; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTNCFQTTPPLWRNPDGQPLCNACGLFLKLHGVRPLSLKTDVIKRRN 50
 ||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 652 CTNCFQTTPPLWRNPDGQPLCNACGLFLKLHGVRPLSLKTDVIKRRN 701

RESULT 2

Q9Y7E8 PRELIMINARY; PRT; 866 AA.
 AC Q9Y7E8;
 DT 01-NOV-1999 (TReMBLrel. 12, Created)
 DT 01-NOV-1999 (TReMBLrel. 12, Last sequence update)
 DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
 DE Major nitrogen regulatory protein.
 GN AREA.
 OS Aspergillus parasiticus.
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
 OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
 OX NCBI_TaxID=5067;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 56775;
 RX MEDLINE=20225467; PubMed=10760588;
 RA Chang P.-K., Yu J., Bhatnagar D., Cleveland T.E.;
 RT "Characterization of Aspergillus parasiticus major nitrogen regulatory
 gene, area";
 RL Biochim. Biophys. Acta 1491:263-266(2000).
 DR EMBL; AF148539; AAD37409.1; -;
 DR HSSP; P17429; 4GAT.
 DR InterPro; IPR000679; Znf_GATA.
 DR Pfam; PF00320; GATA; 1.
 DR PRINTS; PR00619; GATAZNFINGER.
 DR SMART; SM00401; Znf_GATA; 1.
 DR PROSITE; PS00344; GATA_ZN_FINGER_1; 1.
 DR PROSITE; PS50114; GATA_ZN_FINGER_2; 1.
 SQ SEQUENCE 866 AA; 92877 MW; 16FF4D56E823988 CRC64;

Query Match 98.6%; Score 277; DB 3; Length 866;
 Best Local Similarity 98.0%; Pred. No. 8.7e-30;
 Matches 49; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 1 CTNCFTQTTPLRNPDPGLNACGLFLKLGWVRPLSLKTDVIKKNR 50
 Db 664 CTNCFTQTTPLRNPDPGLNACGLFLKLGWVRPLSLKTDVIKKNR 713

RESULT 3

Q9HEW7 PRELIMINARY; PRT; 918 AA.
 AC Q9HEW7;
 DT 01-MAR-2001 (TReMBLrel. 16, Created)
 DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
 DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
 DE Nitrogen response factor NRF1.
 OS Cladosporium fulvum (Fulvia fulva).
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina;
 OC Dothideomycetes et Chaetothyriomycetes incertae sedis;
 OC Mycosphaerellaceae; mitosporic Mycosphaerellaceae; Cladosporium.
 OX NCBI_TaxID=5499;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Perez-Garcia A., Snoeijers S.S., Joosten M.H.A.J., Goosen T.,
 RA De Wit P.J.G.M.;
 RT "Expression of the avirulence gene Avr9 of the fungal tomato pathogen
 Cladosporium fulvum is regulated by the global nitrogen response
 factor NRF1";
 RL Mol. Plant Microbe Interact. 0:0-0(2001).
 DR EMBL; AF312694; AAG48616.1; -;
 DR HSSP; P17429; 4GAT.
 DR InterPro; IPR002965; P-rich_extensn.
 DR InterPro; IPR000679; Znf_GATA.
 DR Pfam; PF00320; GATA; 1.
 DR PRINTS; PR00619; GATAZNFINGER.
 DR PRINTS; PR01217; PRICHEXTENSIN.
 DR SMART; SM00401; Znf_GATA; 1.
 DR PROSITE; PS00344; GATA_ZN_FINGER_1; 1.
 DR PROSITE; PS50114; GATA_ZN_FINGER_2; 1.
 SQ SEQUENCE 918 AA; 99077 MW; E98198D999BCA899 CRC64;

Query Match 98.6%; Score 277; DB 3; Length 918;
 Best Local Similarity 98.0%; Pred. No. 9.3e-30;
 Matches 49; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTNCFTQTTPLRNPDPGLNACGLFLKLGWVRPLSLKTDVIKKNR 50
 Db 673 CTNCFTQTTPLRNPDPGLNACGLFLKLGWVRPLSLKTDVIKKNR 722

RESULT 4

O60043 PRELIMINARY; PRT; 944 AA.
 AC O60043;
 DT 01-AUG-1998 (TReMBLrel. 07, Created)
 DT 01-AUG-1998 (TReMBLrel. 07, Last sequence update)
 DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
 DE Nitrogen response regulator.
 GN NRRL.
 OS Metarhizium anisopliae.
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
 OC Hypocreales; Clavicipitaceae; mitosporic Clavicipitaceae; Metarhizium.
 OX NCBI_TaxID=5530;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Mel;
 RX MEDLINE=99070047; PubMed=9852945;
 RA Screen S.E., Bailey A.M., Charnley K., Cooper R., Clarkson J.M.;
 RT "Isolation of a nitrogen response regulator gene (nr1) from
 Metarhizium anisopliae";
 RL Gene 221:17-24(1998).
 DR EMBL; AJ006468; CAA07052.1; -;
 DR HSSP; P17429; 4GAT.
 DR InterPro; IPR000679; Znf_GATA.
 DR Pfam; PF00320; GATA; 1.
 DR PRINTS; PR00619; GATAZNFINGER.
 DR SMART; SM00401; Znf_GATA; 1.
 DR PROSITE; PS00344; GATA_ZN_FINGER_1; 1.
 DR PROSITE; PS50114; GATA_ZN_FINGER_2; 1.
 SQ SEQUENCE 944 AA; 99687 MW; ALA723E58C23EAO CRC64;

Query Match 98.6%; Score 277; DB 3; Length 944;
 Best Local Similarity 98.0%; Pred. No. 9.6e-30;
 Matches 49; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTNCFTQTTPLRNPDPGLNACGLFLKLGWVRPLSLKTDVIKKNR 50
 Db 663 CTNCFTQTTPLRNPDPGLNACGLFLKLGWVRPLSLKTDVIKKNR 712

RESULT 5

Q9P952 PRELIMINARY; PRT; 1075 AA.
 AC Q9P952;
 DT 01-OCT-2000 (TReMBLrel. 15, Created)
 DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
 DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
 DE Transcription factor ScGATA-6.
 GN SCGT-6.
 OS Schizophyllum commune (Bracket fungus).
 OC Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;
 OC Agaricales; Schizophyllaceae; Schizophyllum.
 OX NCBI_TaxID=5334;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Yamagishi K.;
 RT "putative transcription factor ScGATA-6";
 RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AB043467; BAA96108.1; -;
 DR HSSP; P17429; 4GAT.
 DR InterPro; IPR000679; Znf_GATA.
 DR Pfam; PF00320; GATA; 1.
 DR PRINTS; PR00619; GATAZNFINGER.
 DR SMART; SM00401; Znf_GATA; 1.

```
DR PROSITE; PS00344; GATA_ZN_FINGER_1; 1.
DR PROSITE; PS50114; GATA_ZN_FINGER_2; 1.
SQ SEQUENCE 1075 AA; 113322 MW; D29343211C02FFEA CRC64;
```

Query Match 87.9%; Score 247; DB 3; Length 1075;
Best Local Similarity 88.0%; Pred. No. 1.6e-25;
Matches 44; Conservative 2; Mismatches 4; Indels

Qy 1 C T N C F T Q T T P L W R R N P D G Q P L C N A C G L F L K L H G V V R P L S L K T D V I K K N R 50
 | | | | |
Db 957 C T N C Q T T N T P L W R D P E G Q P L C N A C G L F Y K L H G V V R P L S L K T D V I K K N R 1006

RESULT 6

P87016	PRELIMINARY;	PRT;	327 AA.
ID	P87016		
AC	P87016;		
DT	01-JUL-1997	(TREMBLrel. 04, Created)	
DT	01-JUL-1997	(TREMBLrel. 04, Last sequence update)	
DT	01-JUN-2001	(TREMBLrel. 17, Last annotation update)	
DE	DNA-binding protein.		
GN	SAT1		

OS Zygosaccharomyces rouxii (Candida mogii).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Zygosaccharomycetes.

OX	NCBI_TaxID=4956;
RN	[1]
RP	SEQUENCE FROM N.A.
RA	Ushio K Otsuka Y Yoshitaka C Etsuko I

RT "Cloning of the SAT1 gene concerned with salt tolerance of the
RT *Zygosaccharomyces rouxii*.";

RL J. Ferment. Bioeng. 82:16-21(1996).
DR EMBL; D83211; BAA18948.1; -
DR HSSP; P17429; 4GAT.

DR IncePro; IP0000679; Znf_GATA.
DR Pfam; PF00320; GATA; 1.
DR PRINTS; PR00619; GATAZNFINGER.
DR SMART; SM00401; Znf_GATA; 1.

DR PROSITE; PS50114; GATA_ZN_FINGER_2; 1.
KW DNA-binding.
SQ SEQUENCE 327 AA; 37304 MW; 9D7AF1002807283F CRC64;

Query Match	85.4%	Score 240;	DB 3;	Length 327;
Best Local Similarity	86.0%	Pred. No. 4.3e-25;		
Matches 43; Conservative	3;	Mismatches 4;	Indels 0;	

[illegible]

22 240 C1CCHHRI1FENKRDFTWGNFECNACGUE LKJHG VV RPLSEKTDVIRKKQR 295

RESULT 7

O59842
ID O59842 PRELIMINARY; PRT; 298 AA.
AC O59842;
DTE 01-NOV-1968

DT 01-AUG-1998 (TReMBLrel. 07, Created)
DT 01-AUG-1998 (TReMBLrel. 07, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE GATA transcription factor.

GN NREB.
OS Penicillium chrysogenum.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes; 2

OC: Eucotyledonae; Trichocomaceae; mitosporic Trichocomaceae; Penicill
 OX: Eucotyledonae; Trichocomaceae; mitosporic Trichocomaceae; Penicill
 RN: [1]
 RP: SEQUENCE FROM N A

HAAS H., ANGERMAYR K., ZADRA I.;
Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
EMBL; U96385; AAC09045.1; -.

DR HSSP; P17429; 4GAT.
DR InterPro; IPR000679; znf_GATA.
DR Pfam; PF00320; GATA; 1.

DR PRINTS; PR00619; GATA2NFINGER.
DR SMART; SM00401; ZNF_GATA: 1.
DR POSITE; DS00344; CAT4_ZN_FINGER 1. 1

DR PROSITE: PSS0114; GATA-ZN_FINGER_2; 1.
SQ SEQUENCE 298 AA; 32670 MW; C220D36F4993C590 CRC64;
Query Match 72.6%; Score 204; DB 3; Locath 208;

[illegible]

Db I

RESULT 8	
Q9HEV2	
ID	Q9HEV2
AC	Q9HEV2; PRELIMINARY; PRT; 312 AA.
DT	01-MAR-2001 (TREMBlrel. 16, Created)
DT	01-MAR-2001 (TREMBlrel. 16, Last sequence update)

DI 01-JUN-2002 (TREMBLER, ZI, Last annotation update)
DE GATA factor AREB alpha.
OS AREB.
OS *Emricella nidulans* (Aspergillus nidulans).

OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; Emericella.
OX NCBI_TaxID=5072;

[1]
SEQUENCE FROM N.A.
RX MEDLINE=8918944; PubMed=2651886;
PA Arst H N Tr Tollezou D Coddier M Y.

"A translocation associated, loss-of-function mutation in the nitrogen metabolite repression regulatory gene of *Aspergillus nidulans* can revert intracistronically.";

RL Mol. Gen. Genet. 215:364-367 (1989).
[2]
RP SEQUENCE FROM N. A. SUBV-3 11200110
DY MEDLINE 3120622

RA MEDLINE=212060224; PubMed=11309119;
RA Conlon H., Zadra I., Haas H.N. Jr., Jones M.G., Caddick M.X.;
RA "The Aspergillus nidulans GATA transcription factor gene areB encodes
RT at least three proteins and features three classes of mutation".
RT

KL MOL: Microbio1. 40:361-375 (2001).
DR EMBL: AF320976; AAG49353.1; -.
DR HSSP: P17429; 4CAT.
DR InterPro: IPR000679; Znf_GATA.

```
DR PFAM: PF00320; GATA; 1
DR PRINTS; PR00619; GATAZNFINGER.
DR SMART; SM00401; ZNF_GATA; 1.
```

```
DR DR PROSITE; PS00344; GATA_ZN_FINGER_1; 1.  
DR DR PROSITE; PS50114; GATA_ZN_FINGER_2; 1.  
SQ SEQUENCE 312 AA; 34279 MW; 35A73B7526FE5F88 CRC64;
```

Query Match	70.5%	Score 198;	DB 3;	Length 312;
Best Local Similarity	74.0%;	Pred. No. 2.8e-19;		
Matches 37; Conservative	2;	Mismatches 11;	Indels 0;	Gaps 0;

1 CTNCTQTTPPLWRRRNDGQPLCNACGLFKLHGVVRPLSLKTDVIKKNR 50

DD 17 CQNCGTSKTFPLWKRDELSVCLNACGLFLKJLHGRPRPISUKTDTVJRSNRK 66

RESULT 9

Q9HEV3
ID Q9HEV3 PRELIMINARY; PRT; 320 AA.
AC Q9HEV3; 320

DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE GATA factor APPR beta

GN
OS
OC
AREB.
GN

factor and sex.

EmERICELLA nidulans (Aspergillus nidulans).
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;

OS	Drosophila melanogaster (Fruit fly).
OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC	Ephydroidea; Drosophilidae; Drosophila.
OX	NCBI_TaxID=7227;
RN	[1]
RP	STAPLETON M., BROKSTEIN P., HONG L., AGBAYANI A., CARLSON J.,
RA	CHAMPE M., CHAVEZ C., DORSETT V., DRESNEK D., FARFAN D., FRISE E.,
RA	GEORGE R., GONZALEZ M., GUARIN H., KRONMILLER B., LI P., LIAO G.,
RA	MIRANDA A., MUNGALL C.J., NUNOO J., PACLEB J., PARAGAS V., PARK S.,
RA	PATEL S., PHOUANENAVONG S., WAN K., YU C., LEWIS S.E., RUBIN G.M.,
RA	CELSNIKER S.;
RL	Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
DR	EMBL; AY095095; AAM11423.1; -.
SQ	SEQUENCE 486 AA; 51606 MW; BC0E10E9B8D40A0B CRC64;

Query Match	68.7%	Score 193;	DB 5;	Length 486;
Best Local Similarity	66.0%	Pred. No. 2.2e-18;		
Matches 33;	Conservative 8;	Mismatches 9;	Indels 0;	Gaps 0;

QY	1	CTNCTGTTTTLWRRNP	GGQLCNACGLFLKLHG	VVRPLSLKTDV	IKRNR	50
Db	174	CTNCGTTTTLWRRND	GGVPCNACGLYYLKLHG	VNRPLAMRK	DGIGTQTRK	223

```

RESULT 15
Q9NJ18
ID Q9NJ18 PRELIMINARY; PRT; 324 AA.
AC Q9NJ18;
DC Q9NJ18;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Zinc finger DNA binding protein panner.
GN PNR.
OS Ceratitis capitata (Mediterranean fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Tephritidae; Tephritidae; Ceratitis.
OC NCBI_TaxID=7213;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=KENAEO;
RX MEDLINE=20171265; PubMed=10704387;
RA Wulbeck C., Simpson P.;
RT "Expression of achaete-scute homologues in discrete proneural clusters
RT on the developing notum of the medfly ceratitis capitata, suggests a
RT common origin for the stereotyped bristle patterns of higher
RT Diptera.;"
RL Development 127:1411-1420(2000).
RL [2]
RN RN
RP SEQUENCE FROM N.A.
RC STRAIN=KENAEO;
RX Wulbeck C., Simpson P.;
RA Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases

```

```

RL Submitted (SEP-1999) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AF184154; AAF61481.1; -.
DR HSSP; PJ7679; 1GNF.
DR InterPro; IPR001164; hRIP_like.
DR InterPro; IPR000679; ZnF_GATA.
DR Pfam; PF00320; GATA; 2.
DR PRINTS; PR00619; GATAZNFINGER.
DR SMART; SM00401; ZnF_GATA; 2.
DR PROSITE; PS00344; GATA_ZN_FINGER_1; 2.
DR PROSITE; PS01114; GATA_ZN_FINGER_2; 2.
SQ SEQUENCE      324 AA;  34795 MW;  CFCE5FD9D3CBD290B   CRC64;

Query Match          67.6%; Score 190; DB 5; Length 324;
Best Local Similarity 64.0%; Pred. No. 3.8e-18;
Matches 32; Conservative 9; Mismatches 9; Indels 0; Gaps 0;

QY      1 CTNCFTQTPTPLWRNRDDGOLPCNACGLEFLKLHGCVRPPLSKLTVDIVIKRRN 50
       |||||::||| ||||| :||:|||||: ||||| |||||::: | | : |
Db     165 CTCNGCFRTTTLWRNRNEGPVPCNACGLYPKLLHGVNRPFLAMRKDIOTKRK 214

```

Mon Jan 6 08:17:56 2003

us-09-725-010-3.rspt

Page 6

Search completed: January 3, 2003, 19:50:35
Job time : 49 secs


```

RESULT 7
AREA ASPNG
ID AREA ASPNG STANDARD; PRT; 882 AA.
AC O13412;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Nitrogen regulatory protein area.
GN AREA.
OS Aspergillus niger.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eutiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
OX NCBI_TaxID=5061;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=N402;
RA Maccabe A.P., Vanhanen S.A.S., Sollewijn Gelpke M.,
RA van de Vondervoort P., Arst H.N., Visser J.;
RL Submitted (SEP-1994) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: MAJOR NITROGEN REGULATORY PROTEIN. POSITIVELY ACTING
CC REGULATORY GENE OF NITROGEN METABOLITE REPRESSION.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- SIMILARITY: CONTAINS 1 GATA-TYPE ZINC FINGER.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Y11006; CAA71897.1; -
CC HSSP; P17429; 4GAT.
CC InterPro; IPR000679; Znf_GATA.
CC Pfam; PF00320; GATA; 1.
CC PRINTS; PR00619; GATAZNFINGER.
CC SMART; SM00401; Znf_GATA; 1.
CC PROSITE; PS00344; GATA_ZN_FINGER_1; 1.
CC PROSITE; PS01114; GATA_ZN_FINGER_2; 1.
CC Transcription regulation; Activator; DNA-binding; Zinc-finger;
KW Nuclear protein; Nitrate assimilation.
FT ZN_FING 676 700 GATA-TYPE.
SQ SEQUENCE 882 AA; 94518 MW; 97A502936B94E5AE CRC64;

Query Match 92.5%; Score 258; DB 1; Length 882;
Best Local Similarity 92.0%; Pred. No. 7.le-25;
Matches 46; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 CTNCGTKTTPLRNRRPQGQPLCNACGLFLKLGWVRPLSLKTDVIKKRQR 50
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 676 CTNCFQTTPLRNRRPQGQPLCNACGLFLKLGWVRPLSLKTDVIKKRNR 725

RESULT 8
AREA GIBFU
ID AREA GIBFU STANDARD; PRT; 971 AA.
AC P78688;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Nitrogen regulatory protein area.
GN AREA.
OS Gibberella fujikuroi (Bakanae and foot rot disease fungus) (Fusarium
OS moniliforme).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Hypocreales; Nectriaceae; Gibberella.
OX NCBI_TaxID=5127;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=m567;

Query Match 92.5%; Score 258; DB 1; Length 882;
Best Local Similarity 92.0%; Pred. No. 7.le-25;
Matches 46; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 CTNCGTKTTPLRNRRPQGQPLCNACGLFLKLGWVRPLSLKTDVIKKRQR 50
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 676 CTNCFQTTPLRNRRPQGQPLCNACGLFLKLGWVRPLSLKTDVIKKRNR 725

RESULT 9
NIT2_NEUCR
ID NIT2_NEUCR STANDARD; PRT; 1036 AA.
AC P19212;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Nitrogen catabolic enzyme regulatory protein.
GN NIT-2.
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariales; Sordariaceae; Neurospora.
OX NCBI_TaxID=5141;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=74-OR23-1A;
RA MEDLINE=90158568; PubMed=2137552;
RA Fu Y.-H., Marzluf G.A.;
RT "nit-2, the major nitrogen regulatory gene of Neurospora crassa,"
RT encodes a protein with a putative zinc finger DNA-binding domain."
RT Mol. Cell. Biol. 10:1056-1065(1990).
RN [2]
RP MUTAGENESIS.
RX MEDLINE=91186820; PubMed=2150539;
RA Fu Y.-H., Marzluf G.A.;
RT "Site-directed mutagenesis of the 'zinc finger' DNA-binding domain of
RT the nitrogen-regulatory protein NIT2 of Neurospora."
RL Microbiol. 4:1847-1852(1990).
CC -!- FUNCTION: MAJOR NITROGEN REGULATORY PROTEIN. DURING CONDITIONS
CC OF NITROGEN LIMITATION IT TURNS ON THE EXPRESSION OF GENES FOR
CC ENZYMES WHICH ARE REQUIRED FOR THE USE OF A VARIETY OF SECONDARY
CC NITROGEN SOURCES, INCLUDING NITRATES, PURINES, AMINO ACIDS, AND
CC PROTEINS.

```


RA Mosedale D., Nakahara K., Namath A., Norgren R., Oefner P., Oh C.,
 RA Petel F.X., Roberts D., Sehl P., Schramm S., Shogren T., Smith V.,
 RA Taylor P., Wei Y., Yelton M., Botstein D., Davis R.W.;
 RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: POSITIVE NITROGEN REGULATORY PROTEIN. REQUIRED FOR
 CC THE ACTIVATION OF TRANSCRIPTION OF A NUMBER OF GENES (INCLUDING
 CC THE ALLANTOIN PATHWAY GENES) IN RESPONSE TO THE REPLACEMENT OF
 CC GLUTAMINE BY GLUTAMATE AS SOURCE OF NITROGEN. BINDS THE NITROGEN
 CC UPSTREAM ACTIVATION SEQUENCE OF GLN1. THE GENE ENCODING GLUTAMINE
 CC SYNTHETASE, URE2 MAY CATALYTICALLY INACTIVATE GLN3 IN RESPONSE TO
 CC AN INCREASE IN THE INTRACELLULAR CONCENTRATION OF GLUTAMINE.
 CC -!- SUBCELLULAR LOCATION: Nuclear.
 CC -!- SIMILARITY: CONTAINS 1 GATA-TYPE ZINC FINGER.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; M35267; RAA34645.1; -;
 DR EMBL; U18796; AAB64575.1; -;
 DR PIR; S22280; S22280.
 DR HSP; P17429; 4GAT.
 DR TRANSFAC; T02818; -;
 DR SGD; S0000842; GLN3.
 DR InterPro; IPR000679; Znf_GATA.
 DR Pfam; PF00320; GATA; 1.
 DR PRINTS; PR00619; GATAZNFINGER.
 DR SMART; SM00401; Znf_GATA; 1.
 DR PROSITE; PS00344; GATA_ZN_FINGER_1; 1.
 DR PROSITE; PS01114; GATA_ZN_FINGER_2; 1.
 KW Transcription regulation; Activator; DNA-binding; Zinc-finger;
 KW Nuclear protein; Nitrate assimilation.
 FT ZN_FING 306 330 GATA-TYPE.
 FT DOMAIN 351 361 ARG/LYS-RICH (BASIC).
 FT CONFLICT 474 474 P -> G (IN REF. 1).
 FT SEQUENCE 730 AA; 79382 MW; 3159E184469942E CRC64;
 SQ
 Query Match 72.4%; Score 202; DB 1; Length 730;
 Best Local Similarity 75.0%; Pred. No. 6.8e-18;
 Matches 36; Conservative 4; Mismatches 8; Indels 0; Gaps 0;
 OY 1 CTNCGTKTTPLRNRRPQGLNACGLFLKLGHVVRPLSLKTDVIKRR 48
 Db 306 CFNCKTFTPLWRNRRSPGNTLCNACGLFQKLHGTMRPLSLKSDVIKRR 353
 RESULT 14
 GZF3_YEAST
 ID GZF3_YEAST STANDARD; PRT; 551 AA.
 AC P42944;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE GZF3 protein.
 GN GZF3 OR YJL110C OR J0806.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxID=4932;
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288C / FY1679;
 RX MEDLINE=96090136; PubMed=7403851;
 RA Rasmussen S.W.;
 RT "A 37.5 kb region of yeast chromosome X includes the SME1, MEF2, GSH1
 RT and CSD3 genes, a TCP-1-related gene, an open reading frame similar
 RT to the DAL80 gene, and a tRNA(Arg).";
 RL Yeast 11:873-883(1995).
 RN [2]

RP SEQUENCE FROM N.A.
 RC STRAIN=S288C / FY1679;
 RX MEDLINE=97103775; PubMed=8948101;
 RA Cziepluch C., Kordes E., Pujol A., Jauniaux J.-C.;
 RT "Sequencing analysis of a 40.2 kb fragment of yeast chromosome X
 RT reveals 19 open reading frames including URA2 (5' end), TRK1, PBS2,
 RT SPT10, GCD14, RPE1, PHO86, NCA3, ASF1, COT7, GZF3, two tRNA genes,
 RT three remnant delta elements and a Ty4 transposon.";
 RL Yeast 12:1471-1474(1996).
 CC -!- SUBCELLULAR LOCATION: Nuclear (Potential).
 CC -!- SIMILARITY: CONTAINS 1 GATA-TYPE ZINC FINGER.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; X85021; CAA59384.1; -;
 DR EMBL; Z49385; CAA89405.1; -;
 DR EMBL; X86353; CAA60126.1; -;
 DR HSP; P17429; 4GAT.
 DR TRANSFAC; T02827; -;
 DR SGD; S0003646; GZF3.
 DR InterPro; IPR000679; Znf_GATA.
 DR Pfam; PF00320; GATA; 1.
 DR PRINTS; PR00619; GATAZNFINGER.
 DR SMART; SM00401; Znf_GATA; 1.
 DR PROSITE; PS00344; GATA_ZN_FINGER_1; 1.
 DR PROSITE; PS01114; GATA_ZN_FINGER_2; 1.
 KW Transcription regulation; DNA-binding; Zinc-finger; Nuclear protein.
 FT ZN_FING 131 155 GATA-TYPE.
 FT SEQUENCE 551 AA; 60519 MW; 807243799BFAB704 CRC64;
 SQ
 Query Match 72.0%; Score 201; DB 1; Length 551;
 Best Local Similarity 74.0%; Pred. No. 7e-18;
 Matches 37; Conservative 2; Mismatches 11; Indels 0; Gaps 0;
 OY 1 CPNCGTKTTPLRNRRPQGLNACGLFLKLGHVVRPLSLKTDVIKRR 50
 Db 131 CKNCLSTTTPLRNRRDHGAMLCNACGLFLKLGKPRPISLKTVDIKSRNR 180
 RESULT 15
 PNR_DROME
 ID PNR_DROME STANDARD; PRT; 540 AA.
 AC P52168;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE GATA-binding factor-A (Transcription factor GATA-A) (dGATA-A) (Pannier
 DE protein).
 GN PNR OR GATA-A.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
 OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
 OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=94139551; PubMed=7916677;
 RA Winick J., Abel T., Leonard M.W., Michelson A.M.,
 RA Chardon-Loriaux I., Holmgren R.A., Maniatis T., Engel J.D.;
 RT "A GATA family transcription factor is expressed along the embryonic
 RT dorsoventral axis in Drosophila melanogaster.";
 RL Development 119:1055-1065(1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94139568; PubMed=7916679;
 RA Ramain P., Heitzler P., Haenlin M., Simpson P.;
 RT "Pannier, a negative regulator of achaete and scute in Drosophila,
 RN

QY 1 CTNCGTKTPTLWRRNPQGPCLNACGLFLKLGVVPRPLSLKTDVIKKRQR 50
 |||||:|| ||||| |:|||||||: ||||| |||||: |:|:|:
 Db 226 CTNCGTRTTTLWRRNNDGPEVCNACGLYKLGVNRPPLAWRKDGIQTRKR 275

Search completed: January 3, 2003, 19:48:37
Job time : 15 secs